



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Τεχνικές Εξόρυξης Διεργασιών και Εφαρμογή σε Δεδομένα από τον Τραπεζικό Τομέα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΔΕΣΠΟΙΝΑΣ Α. ΔΕΛΗΠΑΛΛΑ

Επιβλέπων: Δημήτρης Φουσκάκης
Καθηγητής ΕΜΠ

Αθήνα, Σεπτέμβριος 2023



Τεχνικές Εξόρυξης Διεργασιών και Εφαρμογή σε Δεδομένα από τον Τραπεζικό Τομέα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΔΕΣΠΟΙΝΑΣ Α. ΔΕΛΗΠΑΛΛΑ

Επιβλέπων: Δημήτρης Φουσκάκης
Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Δημήτρης Φουσκάκης
Καθηγητής ΕΜΠ

.....
Ιωάννης Κολέτσος
Αναπληρωτής Καθηγητής ΕΜΠ

.....
Βασίλειος Κοκκίνης
Αναπληρωτής Καθηγητής ΕΜΠ



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Copyright © - All rights reserved. Με την επιφύλαξη παντός δικαιώματος.
2023.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

...

Περίληψη

Η Εξόρυξη Διεργασιών αποτελεί έναν σχετικά νέο κλάδο της επιστήμης που συνδυάζει την Στατιστική με την Εξόρυξη Δεδομένων και την Υπολογιστική Νοημοσύνη. Αφετηρία της Εξόρυξης Διεργασιών είναι ένα Αρχείο Καταγραφής Γεγονότων. Κάθε Γεγονός ανήκει σε μία Υπόθεση, ανταποκρίνεται σε μία Δραστηριότητα της Διεργασίας και περιέχει μία χρονική σήμανση. Μέσα από αυτό το αρχείο είναι κανείς σε θέση να ανακαλύψει τον τρόπο με τον οποίο εκτελείται η υπό μελέτη Διεργασία στην πραγματικότητα. Αρχικά μπορεί κανείς κατά την Ανακάλυψη του Μοντέλου να εξάγει ένα Μοντέλο σε μορφή Διαγράμματος Ροής που να περιγράφει την Διεργασία του Αρχείου Καταγραφής. Έπειτα, στον Έλεγχο Προσαρμογής μπορεί να ελέγξει τον βαθμό στον οποίο ένα σχεδιασμένο μοντέλο αποκλίνει από την πραγματικότητα ή τον βαθμό στον οποίο διαφέρουν δύο μοντέλα μεταξύ τους. Τέλος στην Ανάλυση Βασικών Αιτιών μπορεί κανείς να μελετήσει τις αιτίες οι οποίες οδηγούν σε δυσλειτουργίες της Διεργασίας, όπως καθυστερήσεις Υποθέσεων ή Υποθέσεις που δεν ανταποκρίνονται στην συνήθη συμπεριφορά του Μοντέλου.

Σκοπός της διπλωματικής εργασίας είναι η παρουσίαση των παραπάνω εννοιών και των κυριότερων μεθόδων που έχουν αναπτυχθεί για κάθε μία από αυτές και τέλος η εφαρμογή τεχνικών Εξόρυξης Διεργασιών με την βοήθεια του εργαλείου QPR Process Analyzer σε ένα Αρχείο Καταγραφής Γεγονότων που περιέχει δεδομένα για την διαδικασία απόδοσης δανείων σε μία τράπεζα.

Λέξεις Κλειδιά

Εξόρυξη Διεργασιών, Αρχείο Καταγραφής Γεγονότων, Ανακάλυψη του Μοντέλου, Έλεγχος Προσαρμογής, Ανάλυση Βασικών Αιτιών, Διάγραμμα Ροής

Abstract

Process Mining is a relatively new branch of science that combines Statistics with Data Mining and Computational Intelligence. The starting point of Process Mining is an Event Log. Each Event belongs to a Case, corresponds to a Process Activity and contains a time label. Through this record one is able to discover how the Process under study is actually performed. Initially one can during Model Discovery extract a Model in the form of a Flowchart describing the Process of the Log File. Then, in the Conformance Checking one can check the extent to which a designed model deviates from reality or the extent to which two models differ from each other. Finally, in the Root Causes Analysis one can study the causes that lead to Process bottlenecks, such as Case delays or Cases that do not correspond to the normal behavior of the Model.

The purpose of this thesis is to present the above concepts and the main methods developed for each of them and finally to apply Process Extraction techniques with the help of the QPR Process Analyzer tool to an Event Log containing data on the loan granting process in a bank.

Keywords

Process Mining, Event Log, Model Discovery, Conformance Checking, Root Causes Analysis, Flowchart

στην Δήμητρα, τον Χρήστο και τον Στέργιο

Ευχαριστίες

Θα ήθελα καταρχάς να ευχαριστήσω τον καθηγητή κ. Δημήτρη Φουσκάκη για την επίβλεψη αυτής της διπλωματικής εργασίας. Επίσης ευχαριστώ ιδιαίτερα την ITMC και τον κ. Γιάννη Χαραλάμπους για την παροχή του QPR Process Analyzer και για το έναυσμα να ασχοληθώ με το αντικείμενο της Εξόρυξης Διεργασιών. Θα ήθελα ακόμα να ευχαριστήσω την οικογένεια μου για την ηθική και υλική υποστήριξη καθ' όλα τα χρόνια των σπουδών μου. Τέλος θέλω να ευχαριστήσω τους φίλους και την μεγάλη μου οικογένεια που υπήρξαν έμπνευση και κινητήριοις δύναμη για να βάζω τον πήχη κάθε φορά ψηλότερα.

Περιεχόμενα

Περίληψη	1
Abstract	2
Ευχαριστίες	4
1 Εισαγωγή στο Αντικείμενο	10
1.1 Αντικείμενο της διπλωματικής	12
2 Αναλυτική Ανασκόπηση Θεωρίας	13
2.1 Η Ανακάλυψη του Μοντέλου	13
2.1.1 Ευρετικός Αλγόριθμος Εξόρυξης (Heuristic Miner)	21
2.1.2 Γενετικός Αλγόριθμος Εξόρυξης (Genetic Miner)	25
2.1.3 Ασαφής Αλγόριθμος Εξόρυξης (Fuzzy Miner)	29
2.2 Η Προσαρμογή του Μοντέλου	30
2.2.1 Έλεγχος Προσαρμογής μεταξύ ενός μοντέλου και ενός Αρχείου Καταγραφής Γεγονότων	30
2.2.2 Έλεγχος Προσαρμογής μεταξύ δύο μοντέλων	32
2.3 Η Ανάλυση των Βασικών Αιτίων	33
2.3.1 Η Μέθοδος του Hompes	34
2.3.2 Η Μέθοδος των Van der Aalst και Qafari	36
2.3.3 Ο Αλγόριθμος ΑΙΠΙΑ - PM	38
3 Εφαρμογή των Τεχνικών Εξόρυξης Διεργασιών σε Δεδομένα από τον Τραπεζικό Τομέα	40
3.1 Τα Δεδομένα του Προβλήματος	40
3.2 Εργαλείο ProM	42
3.3 Εργαλείο QPR Process Analyzer	42
4 Προβολή Δεδομένων και Ανάλυση	44
4.1 Το Αρχείο Καταγραφής Γεγονότων	44
4.2 Εφαρμογή Μεθόδων Εξόρυξης Διεργασιών	45
4.2.1 Ανακάλυψη του Μοντέλου της Διεργασίας	45
4.2.2 Έλεγχος Προσαρμογής	50
4.3 Εξόρυξη Διεργασιών σε Επιχειρησιακά Περιβάλλοντα	52
4.3.1 Το Αρχείο Καταγραφής Υποθέσεων	53

4.3.2 Αναφορές για τα Χαρακτηριστικά των Υποθέσεων	56
4.3.3 Ανάλυση Βασικών Αιτιών για Υποθέσεις με Καθυστέρηση	61
5 Συμπεράσματα - Μελλοντική Έρευνα	63
Παραρτήματα	64
Α΄ Κώδικες του QPR Process Analyzer	65
Βιβλιογραφία	79
Συντομογραφίες - Αρκτικόλεξα - Ακρωνύμια	80
Απόδοση ξενόγλωσσων όρων	81

Κατάλογος Εικόνων

2.1	Αποδεκτό Δίκτυο Petri	20
2.2	Γράφημα Άμεσης Ακολουθίας για το Αρχείο Καταγραφής Γεγονότων L_1	20
2.3	Δίκτυο Petri για το Αρχείο Καταγραφής Γεγονότων L_1	21
2.4	Δίκτυο Petri με σύζευξη και διάζευξη	24
2.5	Απεικόνιση του Αιτιολογικού Πίνακα σε Δίκτυο Petri	27
2.6	Απεικόνιση μίας Διεργασίας τύπου σπαγκέτι. [1]	29
3.1	Αρχείο Καταγραφής Γεγονότων	41
3.2	Αρχείο Καταγραφής Υποθέσεων	41
4.1	Αρχείο Καταγραφής Γεγονότων	44
4.2	Συνοπτική Επισκόπηση του Αρχείου Καταγραφής στο περιβάλλον του ProM .	47
4.3	Συνοπτική Επισκόπηση του Αρχείου Καταγραφής στο περιβάλλον του ProM .	47
4.4	Απεικόνιση του Μοντέλου ως Γράφημα Άμεσης Ακολουθίας	48
4.5	Η εξαγωγή του Μοντέλου με βάση τον Ευρετικό Αλγόριθμο	49
4.6	Ευρετικός Αλγόριθμος με πρόσθετους περιορισμούς	50
4.7	Δίκτυο Petri για το Μοντέλο M_1	50
4.8	Έλεγχος Προσαρμογής μεταξύ του Αρχείου Καταγραφής Γεγονότων και του Μοντέλου M_1	51
4.9	Έλεγχος Προσαρμογής μεταξύ του Αρχείου Καταγραφής Γεγονότων και του Μοντέλου M_1 (Αναλυτικά Αποτελέσματα)	52
4.10	Αναφορά στο QPR Process Analyzer σχετικά με τα Γεγονότα Έναρξης και Λήξης και τα Ίχνη της Διαδικασίας	53
4.11	Αρχείο Καταγραφής Υποθέσεων	54
4.12	Αναφορά με βάση την Κατάσταση του Δανείου	57
4.13	Πίνακας της Αναφοράς με βάση την Κατάσταση του Δανείου	58
4.14	Παράδειγμα επιλογής φίλτρου	58
4.15	Η Αναφορά έχοντας φιλτράρει τα δεδομένα με βάση τον Μήνα Καταχώρησης και Την Επιχειρηματική Μονάδα	59
4.16	Αναφορά για τις Υποθέσεις που αφορούν Δάνεια που έχουν απορριφθεί	59
4.17	Πίνακας της Αναφοράς για τα Δάνεια που έχουν απορριφθεί	60
4.18	Αναφορά για την Διάρκεια των Υποθέσεων (μέρος 1)	60
4.19	Αναφορά για την Διάρκεια των Υποθέσεων (μέρος 2)	61
4.20	Αναφορά για την Διάρκεια των Υποθέσεων (με φίλτρο διάρκειας 11 ως 46 ημέρες)	61

4.21 Αναφορά για τους παράγοντες που σχετίζονται με την καθυστέρηση μίας Υ- πόθεσης	62
--	----

Κατάλογος Πινάκων

2.1	Αρχείο Καταγραφής Γεγονότων L	15
2.2	Αρχείο Καταγραφής Γεγονότων της Εταιρείας	16
2.3	Αξιολόγηση της ποιότητας των Αρχείων Καταγραφής Γεγονότων κατά τον van der Aalst	18
2.4	Απτιολογικός Πίνακας	26
2.5	Αρχείο Καταγραφής Γεγονότων της Υπηρεσίας	37

Εισαγωγή στο Αντικείμενο

Στην εποχή μας έχει καθιερωθεί η χρήση της τεχνολογίας στους περισσότερους τομείς της ζωής. Δραστηριότητες που εκτείνονται από την έκδοση εγγράφων και την σύναψη συμβολαίων μέχρι την ψυχαγωγία και την ενημέρωση πραγματοποιούνται μέσω εφαρμογών, ιστοσελίδων και ειδικών λογισμικών. Αυτό από την μία κάνει πιο εύκολη την καθημερινότητα πολλών ανθρώπων και από την άλλη συγκεντρώνει έναν μεγάλο όγκο δεδομένων προς αξιοποίηση. Οι μεγάλοι όγκοι φακέλων και χαρτιών που υπήρχαν άλλοτε σε αποθήκες υπηρεσιών πλέον δίνουν την θέση τους σε μεγάλα αρχεία αποθηκευμένα σε κάποιον server. Ο όγκος και η πολυπλοκότητα των δεδομένων αυτών είναι τέτοια, ώστε οι παραδοσιακές μέθοδοι στατιστικής δεν αρκούν πια για την ανάλυση τους και έτσι γεννιέται η ανάγκη ανάπτυξης μεθόδων που είναι πιο “έξυπνες” συνδυάζοντας την στατιστική με την τεχνητή νοημοσύνη.

Στον ωκεανό των δεδομένων που υπάρχουν διαθέσιμα, πέρα από την εξαγωγή συμπερασμάτων για μεμονωμένα μεγέθη, συχνά μπορεί κανείς να ανακαλύψει δραστηριότητες, βήματα που διαδέχονται το ένα το άλλο σχηματίζοντας μονοπάτια και έτσι να εξάγει συμπεράσματα που αφορούν συμπεριφορές ή τον τρόπο με τον οποίο λειτουργούν ορισμένες διεργασίες. Τότε γίνεται λόγος για Εξόρυξη Διεργασιών, δηλαδή για την διαδικασία με την οποία μέσα από ένα σύνολο Δεδομένων που αφορούν σε μεμονωμένες Δραστηριότητες εξάγονται συμπεράσματα που αφορούν μία Διεργασία, δηλαδή μία ακολουθία βημάτων.

Ορισμός 1.1 (Εξόρυξη Διεργασιών). *Η Εξόρυξη Διεργασιών (process mining) είναι ο τομέας της επιστήμης που ασχολείται με την εξαγωγή γνώσης από Αρχεία Καταγραφής Γεγονότων (event logs), διαθέσιμα από πληροφοριακά συστήματα. Στόχος είναι η ανακάλυψη, ο έλεγχος και η βελτίωση διεργασιών που συμβαίνουν σε πραγματικό χρόνο. Η Εξόρυξη Διεργασιών συνδυάζει την υπολογιστική νοημοσύνη (computational intelligence) με την εξόρυξη δεδομένων (data mining) και την μοντελοποίηση και ανάλυση διεργασιών (process modeling and analysis). [2] Η ουσία του όρου Εξόρυξη Διεργασιών συνοψίζεται στην μέθοδο με την οποία μέσα από ένα σύνολο χρονικά διατεταγμένων δεδομένων που αφορούν την εκτέλεση εργασιών αποκρυσταλλώνεται η περιγραφή μιας διαδικασίας.[3]*

Οι βασικές οπτικές της Εξόρυξης Διεργασιών είναι η Ανακάλυψη (discover), ο Έλεγχος Προσαρμογής (conformance checking) και η Ενίσχυση (enhancement). Η Ανακάλυψη αφορά στην εξαγωγή ενός μοντέλου από δεδομένα που παρέχονται από ένα Αρχείο Καταγραφής Γεγονότων χωρίς χρήση πρότερης γνώσης. Ο Έλεγχος αφορά στην σύγκριση ενός υπάρχοντος μοντέλου με την πραγματικότητα. Η Ενίσχυση αφορά στην βελτίωση του μοντέλου μιας

διεργασίας χρησιμοποιώντας συμπεράσματα από τον τρόπο που λειτουργεί η Διαδικασία. Τέλος, μία άλλη οπτική είναι η Ανάλυση Βασικών Αιτιών (Root Causes Analysis), η οποία αφορά τον εντοπισμό των αιτιών που προκαλούν κωλύματα στην Διαδικασία. [2], [4]

Μέσω της Εξόρυξης Διεργασιών μπορεί κανείς να έχει μια σαφή εικόνα για τον τρόπο με τον οποίο πραγματοποιείται μία Διεργασία στην πραγματικότητα, όταν έχει στην διάθεσή του Δεδομένα από εκτελέσεις συγκεκριμένων ενεργειών που αποτελούν βήματα της Διεργασίας αυτής. Αυτό καθιστά δυνατό τον ανασχεδιασμό μιας Διεργασίας με βάση την πραγματικότητα, τον εντοπισμό κωλυμάτων και δυσλειτουργιών, αλλά και τον εντοπισμό των αιτιών οι οποίες είναι υπεύθυνες για αυτά. Είναι λοιπόν φανερό ότι κάτι τέτοιο μπορεί να έχει εφαρμογή σε μία πληθώρα τομέων. Σύμφωνα με την AIMultiple, μία εταιρεία που δραστηριοποιείται στην ανάλυση και αξιολόγηση τεχνολογικών εργαλείων για εταιρείες [5], οι κυριότεροι τομείς στους οποίους χρησιμοποιείται η Εξόρυξη Διεργασιών είναι οι εξής[6] :

- Διαχείριση Υπηρεσιών Τεχνολογίας (IT Service Management), όπου χρησιμοποιείται με σκοπό να εντοπιστούν επαναλαμβανόμενες ενέργειες που μπορούν να αυτοματοποιηθούν, να εντοπιστούν οι βέλτιστες / συντομότερες διαδρομές για την λύση ενός προβλήματος, να εντοπιστούν οι αιτίες καθυστέρησης κ.α. ώστε να επιτευχθεί η καλύτερη δυνατή παροχή υπηρεσιών τεχνολογίας.
- Τραπεζικός Τομέας, όπου χρησιμοποιείται με σκοπό να βελτιστοποιηθούν Διεργασίες όπως αυτές που αφορούν την Έγκριση ενός Δανείου, την έκδοση μιας κάρτας κ.α., οι οποίες συχνά περιέχουν μεγάλη γραφειοκρατία και μπορούν να γίνουν ιδιαίτερα περίπλοκες
- Λογιστικός Έλεγχος (Audit), όπου χρησιμοποιείται με σκοπό να διαπιστωθεί κατά πόσο η υλοποίηση των αλλαγών που αφορούν μια Διεργασία συμβάλλουν όντως προς την βελτίωσή της , αλλά και για τον εντοπισμό ρίσκων
- Υπηρεσίες Εξυπηρέτησης Πελατών, όπου χρησιμοποιείται με σκοπό να καταγραφεί η εμπειρία του πελάτη και τα συνήθη προβλήματα που αυτός αντιμετωπίζει
- Υγειονομική Περίθαλψη, όπου χρησιμοποιείται με σκοπό να βελτιστοποιηθούν οι διοικητικές διεργασίες, αλλά και να τυποποιηθούν πρακτικές υγειονομικής περίθαλψης και να εντοπιστούν ζητήματα που μπορεί να οδηγήσουν σε λανθασμένες θεραπείες ή καθυστερήσεις, οι οποίες είναι ζωτικής σημασίας για πολλούς ασθενείς

Είναι σαφές ότι όσο η χρήση πληροφοριακών συστημάτων διευρύνεται και σε άλλους τομείς, τόσο θα μπορεί και να επεκτείνεται και η χρήση της Εξόρυξης Διεργασιών. Οι κυριότεροι σκοποί για τους οποίους χρησιμοποιούνται σήμερα πληροφοριακά συστήματα, τα οποία μπορούν να παρέχουν δεδομένα προς επεξεργασία είναι οι εξής: Διαχείριση εργατικού δυναμικού (Workforce Management (WFM)), Διαχείριση Πελατειακών Σχέσεων (Customer Relationship Management (CRM)), Διαχείριση Εφοδιαστικής Αλυσίδας (Supply Chain Management (SCM)), Σχεδιασμό Εταιρικών Πόρων (Enterprise Resource Planning (ERP)) κ.α. [3]. Τα συστήματα αυτά δεν χρησιμοποιούνται με κύριο σκοπό την παροχή δεδομένων για την λειτουργία του Οργανισμού που τα χρησιμοποιεί, αλλά για την διευκόλυνση της λειτουργίας του.

1.1 Αντικείμενο της διπλωματικής

Η παρούσα Διπλωματική Εργασία ασχολείται με την παρουσίαση και ανάλυση των βασικών πλευρών της Εξόρυξης Διεργασιών, με σκοπό να έρθει ο αναγνώστης σε μία πρώτη επαφή με τις θεμελιώδεις έννοιες του αντικειμένου και να γίνει ταυτόχρονα σαφής η λογική που ακολουθείται. Έπειτα στα πλαίσια της Εργασίας αναλύονται Δεδομένα που αφορούν την διαδικασία έγκρισης δανείου μίας τράπεζας. Τα Δεδομένα αυτά αφορούν Δραστηριότητες που εκτελέστηκαν σε διάφορα υποκαταστήματα της τράπεζας, από διαφορετικούς υπαλλήλους σε διαφορετικές χρονικές στιγμές, αποτελούν δηλαδή ένα Αρχείο Καταγραφής Γεγονότων και σκοπός είναι μέσα από αυτά χρησιμοποιώντας την Εξόρυξη Διεργασιών να αντληθεί μία σαφής εικόνα για τον τρόπο που λειτουργεί η διαδικασία έγκρισης δανείων στην τράπεζα αυτή.

Κεφάλαιο 2

Αναλυτική Ανασκόπηση Θεωρίας

Στο κεφάλαιο αυτό παρουσιάζονται αναλυτικά βασικές μεθοδολογίες για την Ανακάλυψη του Μοντέλου, τον Έλεγχο Προσαρμογής και την Ανάλυση Βασικών Αιτιών, που αποτελούν διαφορετικές πλευρές της Εξόρυξης Διεργασιών. Σκοπός του κεφαλαίου είναι η εξοικείωση με τις βασικές έννοιες της Εξόρυξης Διεργασιών και τις βασικές μεθοδολογίες που ακολουθούνται.

2.1 Η Ανακάλυψη του Μοντέλου

Αφετηρία της Εξόρυξης Διεργασιών είναι τα Γεγονότα. Μέσα από τα Γεγονότα εξάγεται το Μοντέλο της Διεργασίας και πραγματοποιούνται όλες οι αναλύσεις. Συνεπώς είναι πολύ σημαντικό να διευκρινιστεί τι θεωρείται Γεγονός. Είναι ευνόητο ότι ένα Γεγονός πρέπει να αναφέρεται σε κάποια Δραστηριότητα που συνέβη. Ωστόσο αυτό δεν είναι αρκετό. Χρειάζεται επιπλέον να υπάρχει πληροφορία για την χρονική στιγμή στην οποία εκτελέστηκε το Γεγονός αυτό, ώστε να μπορούν τα Γεγονότα να διαταχθούν σε χρονολογική σειρά και έτσι να εξαχθούν συμπεράσματα για την σειρά με την οποία η μία Δραστηριότητα διαδέχεται την άλλη σε κάποια Διεργασία, ενώ επιπλέον χρειάζεται να μπορούν τα Γεγονότα να ομαδοποιηθούν με βάση τις εκτελέσεις της Διεργασίας στις οποίες αυτά ανήκουν. Παραδείγματος χάριν εάν έχουμε την πληροφορία ότι η Δραστηριότητα Α συνέβη τις χρονικές στιγμές t_1 και t_2 , η Δραστηριότητα Β συνέβη τις χρονικές στιγμές t_3 και t_4 και η Δραστηριότητα Γ συνέβη την χρονική στιγμή t_5 με $t_1 < t_2 < t_3 < t_4 < t_5$ τότε δεν μπορούμε να έχουμε μία σαφή εικόνα της Διεργασίας καθώς δεν ξέρουμε αν οι εκτελέσεις των Δραστηριοτήτων Α, Β, Γ αφορούν μία ή περισσότερες περιπτώσεις εκτέλεσης της Διεργασίας. Αν επιπλέον προστεθεί η πληροφορία ότι η Δραστηριότητα Α συνέβη την χρονική στιγμή t_1 και αφορούσε την περίπτωση 1, την χρονική στιγμή t_2 και αφορούσε την περίπτωση 2, η Δραστηριότητα Β συνέβη την χρονική στιγμή t_3 και αφορούσε την περίπτωση 2 και την χρονική στιγμή t_4 και αφορούσε την περίπτωση 3 και τέλος ότι η Δραστηριότητα Γ συνέβη την χρονική στιγμή t_5 και αφορούσε την περίπτωση 2, με $t_1 < t_2 < t_3 < t_4 < t_5$, τότε ξέρουμε ότι δύο εκτελέσεις της Διεργασίας είναι η $A \rightarrow B$ και η $A \rightarrow B \rightarrow \Gamma$. Με βάση αυτά ορίζουμε την Θεμελιώδη Συνθήκη Γεγονότων.

Ορισμός 2.2 (Θεμελιώδης Συνθήκη Γεγονότων). *Κάθε Γεγονός πρέπει να ανταποκρίνεται σε μία Δραστηριότητα (activity), δηλαδή ένα καλά ορισμένο βήμα μιας Διεργασίας (process), να σχετίζεται με μία Υπόθεση (case), δηλαδή ένα στιγμιότυπο της Διεργασίας (process instance)*

και να διαθέτει μία Χρονική Σήμανση (timestamp). Επιπλέον ένα γεγονός μπορεί να διαθέτει και περαιτέρω Χαρακτηριστικά (attributes), όπως ο τόπος τους γεγονότος, οι εμπλεκόμενοι κ.α. [7] [8]

Από τον Ορισμό 2.2 προκύπτει ότι Γεγονός θεωρείται μία πλειάδα χαρακτηριστικών, στα οποία πρέπει υποχρεωτικά να περιλαμβάνονται το Αναγνωριστικό της Υπόθεσης με την οποία σχετίζεται το Γεγονός, το όνομα της Δραστηριότητας στην οποία ανταποκρίνεται το Γεγονός και η Χρονική Σήμανση η οποία δείχνει την ακριβή στιγμή που συνέβη (ή ξεκίνησε) το Γεγονός, ενώ πέρα από αυτά ένα Γεγονός ενδέχεται να περιλαμβάνει και άλλα χαρακτηριστικά. Είναι επίσης σαφές ότι πολλά Γεγονότα, χρονικά διατεταγμένα συνθέτουν μία Υπόθεση.

Ορισμός 2.3 (Συνθήκη Υπόθεσης). *Μία Υπόθεση (case) παριστάνεται ως μία ακολουθία γεγονότων, την οποία ονομάζουμε Ίχνος (trace). Δύο ή περισσότερες διαφορετικές Υποθέσεις μπορεί να έχουν το ίδιο Ίχνος.*[8]

Αφού ορίστηκε τι είναι ένα Γεγονός και μία Υπόθεση, τα οποία αποτελούν τις δύο βασικές έννοιες αυτής της Διπλωματικής Εργασίας, είναι χρήσιμο να γίνει κατανοητό με ποιον τρόπο πρέπει να έχει κανείς σε διάθεσή του τα Γεγονότα ώστε να μπορεί να τα επεξεργαστεί. Ο τρόπος αυτός είναι το Αρχείο Καταγραφής Γεγονότων, δηλαδή μία βάση Δεδομένων που περιέχει Γεγονότα.

Ορισμός 2.4 (Αρχείο Καταγραφής Γεγονότων). *Το U_{act} είναι ένα σύνολο από ονομασίες Δραστηριοτήτων και το Ίχνος $\sigma = \langle a_1, a_2, \dots, a_n \rangle \in U_{act}^*$ ¹ είναι μία ακολουθία Δραστηριοτήτων. Ένα πολυσύνολο² Ιχνών $L \in \mathcal{B}(U_{act}^*)$ ³ ονομάζεται Αρχείο Καταγραφής Γεγονότων.*[8]

Αυτό που περιγράφει ο Ορισμός 2.4 είναι ότι ένα Αρχείο Καταγραφής περιέχει Ίχνη Υποθέσεων, δηλαδή πεπερασμένες ακολουθίες από χρονικά διατεταγμένα Γεγονότα, κάθε ένα από τα οποία (Ίχνη) αναπαριστά μία Υπόθεση. Διαφορετικές Υποθέσεις μπορεί να έχουν το ίδιο Ίχνος, δηλαδή την ίδια ακολουθία Δραστηριοτήτων. Ένα Ίχνος παριστάνεται ως μία ακολουθία Δραστηριοτήτων ως εξής: $\langle \text{Δραστηριότητα}_1, \text{Δραστηριότητα}_2, \text{Δραστηριότητα}_3 \rangle$, ενώ ο συμβολισμός $\langle \text{Δραστηριότητα}_1, \text{Δραστηριότητα}_2, \text{Δραστηριότητα}_3 \rangle^k$ υποδηλώνει ότι το Ίχνος $\langle \text{Δραστηριότητα}_1, \text{Δραστηριότητα}_2, \text{Δραστηριότητα}_3 \rangle$ εμφανίζεται k φορές στο Αρχείο Καταγραφής Γεγονότων, δηλαδή ανταποκρίνεται σε k διαφορετικές Υποθέσεις. Από τα παραπάνω συνεπάγεται ότι το Αρχείο Καταγραφής Γεγονότων θα δίνεται ως ένα πολυσύνολο Ιχνών $L = \{tr_1^m, tr_2^n, tr_3^p, \dots\}$, όπου tr_i ένα Ίχνος και m, n, p ο αριθμός εμφανίσεων του εκάστοτε Ίχνους. Ωστόσο, στην πραγματικότητα όταν αναφερόμαστε σε ένα Αρχείο Καταγραφής Γεγονότων, πρόκειται για ένα σύνολο Γεγονότων από το οποίο προκύπτει άμεσα το πολυσύνολο των Ιχνών (και όχι το ίδιο το πολυσύνολο). Αυτό μπορεί να γίνει πιο εύκολα αντιληπτό με ένα παράδειγμα

¹Με U_{act}^* συμβολίζεται το σύνολο όλων των πεπερασμένων ακολουθιών από ονομασίες γεγονότων του U_{act}

²Πολυσύνολο ονομάζεται μία μη διατεταγμένη συλλογή στοιχείων, όπου ένα στοιχείο μπορεί να εμφανίζεται ως μέλος περισσότερες από μία φορές. Π.χ. το πολυσύνολο που δηλώνεται με $\{a, a, a, b, b\}$ είναι το πολυσύνολο που περιέχει 3 φορές το στοιχείο a και 2 φορές το στοιχείο b. [9]

³Με $\mathcal{B}(U_{act}^*)$ συμβολίζεται η βάση του τοπολογικού χώρου U_{act}^* . Μια βάση είναι μια συλλογή ανοικτών συνόλων σε έναν τοπολογικό χώρο που μπορεί να χρησιμοποιηθεί για να δημιουργήσει όλα τα άλλα ανοικτά σύνολα του χώρου.

Παράδειγμα 2.1. Δίνεται ο Πίνακας 2.1, ο οποίος έχει 15 γραμμές και 3 στήλες. Κάθε γραμμή του πίνακα αναπαριστά ένα Γεγονός και κάθε στήλη ένα χαρακτηριστικό. Η πρώτη στήλη μας δίνει πληροφορία για την Υπόθεση στην οποία ανήκει κάποιο Γεγονός, η δεύτερη στήλη μας δίνει πληροφορία για την Δραστηριότητα στην οποία ανταποκρίνεται το Γεγονός και η τρίτη στήλη μας πληροφορεί για την χρονική στιγμή την οποία συνέβη το κάθε Γεγονός. Έτσι για παράδειγμα μπορούμε με βάση την πρώτη γραμμή του Πίνακα 2.1 να δούμε ότι η δραστηριότητα α συνέβη στα πλαίσια της Υπόθεσης 1 στις 21/4/2023, 11:34 κοκ. Συμπεραίνουμε επίσης ότι ο Πίνακας 2.1 περιέχει 15 Γεγονότα, τα οποία ανήκουν σε 4 διαφορετικές Υποθέσεις (Υπόθεση 1, Υπόθεση 2, Υπόθεση 3, Υπόθεση 4 με βάση την πρώτη στήλη) και ανταποκρίνονται σε 4 διαφορετικές Δραστηριότητες (α, β, c, d με βάση την δεύτερη στήλη).

Κωδικός Υπόθεσης	Δραστηριότητα	Χρονική Σήμανση
Υπόθεση 1	a	21-04-2023:11.34
Υπόθεση 2	a	21-04-2023:12.34
Υπόθεση 1	b	21-04-2023:12.40
Υπόθεση 3	a	21-04-2023:13.34
Υπόθεση 2	b	21-04-2023:14.34
Υπόθεση 4	a	22-04-2023:10.30
Υπόθεση 3	c	22-04-2023:11.34
Υπόθεση 3	d	22-04-2023:13.00
Υπόθεση 1	c	22-04-2023:13.34
Υπόθεση 4	c	22-04-2023:14.04
Υπόθεση 2	c	30-04-2023:12.05
Υπόθεση 1	d	30-04-2023:12.09
Υπόθεση 4	b	30-04-2023:13.14
Υπόθεση 2	d	30-04-2023:13.20
Υπόθεση 3	b	30-04-2023:14.30

Πίνακας 2.1: Αρχείο Καταγραφής Γεγονότων L

Κατατάσσοντας τα γεγονότα χρονολογικά ανά Υπόθεση προκύπτει ότι οι 4 Υποθέσεις του Πίνακα είναι οι :

Υπόθεση 1 με Ίχνος $\langle a, b, c, d \rangle$

Υπόθεση 2 με Ίχνος $\langle a, b, c, d \rangle$

Υπόθεση 3 με Ίχνος $\langle a, c, b, d \rangle$

Υπόθεση 4 με Ίχνος $\langle a, c, d \rangle$

Συνεπώς ο Πίνακας 2.1 ισοδυναμεί με το πολυσύνολο Ιχνών $L = \left[\langle a, b, c, d \rangle^2, \langle a, c, d, b \rangle, \langle a, c, b \rangle \right]$.

Από το παράδειγμα 2.1 γίνεται κατανοητό ότι η παρουσίαση του Αρχείου Καταγραφής Γεγονότων με μορφή πίνακα είναι ισοδύναμη με την παρουσίαση του Αρχείου Καταγραφής Γεγονότων ως πολυσύνολο Ιχνών. Καθώς στον πραγματικό κόσμο όταν η Εξόρυξη Διεργασιών έχει κατά κανόνα ως αφετηρία την συλλογή ενός όγκου Γεγονότων, στο εξής ως Αρχείο Καταγραφής Γεγονότων θα νοείται ο Πίνακας, ο οποίος περιέχει το σύνολο των καταγεγραμμένων

Γεγονότων με τρόπο ώστε κάθε γραμμή να αναπαριστά ένα μοναδικό Γεγονός και κάθε στήλη να αναπαριστά ένα χαρακτηριστικό.

Επιπλέον στο παράδειγμα 2.1 παρατηρούμε ότι οι διαφορετικές Υποθέσεις δεν περιέχουν απαραίτητα τον ίδιο αριθμό Γεγονότων.

Παράδειγμα 2.2. *Μία εταιρεία ασχολείται με την έκδοση ορισμένων εγγράφων. Προκειμένου κανείς να παραλάβει το έγγραφο που χρειάζεται πρέπει να υποβάλει μία αίτηση. Η στιγμή υποβολής της αίτησης καταγράφεται στο ηλεκτρονικό σύστημα της εταιρείας. Στη συνέχεια η αίτηση ενδέχεται να γίνει δεκτή ή να απορριφθεί, ενώ ο υπάλληλος που ασχολείται με το εν λόγω βήμα δηλώνει την απόφαση στο ηλεκτρονικό σύστημα, το οποίο καταγράφει την ακριβή ώρα δήλωσης. Έπειτα οι υπάλληλοι αναζητούν το έγγραφο στο ηλεκτρονικό αρχείο της εταιρείας, το οποίο δείχνει εάν αυτό υπάρχει ή όχι και καταγράφει το αποτέλεσμα και την αντίστοιχη ακριβή ώρα αναζήτησης. Τέλος, ο αιτών ενημερώνεται για την έκβαση της αίτησής του μέσω email, όπου φαίνεται η ακριβής ώρα της ενημέρωσης.*

Συγκεντρώνουμε τα δεδομένα που προκύπτουν και έχουμε τον Πίνακα 2.2.

Όνομα Πελάτη	Δραστηριότητα	Χρονική Σήμανση	Υπάλληλος
A	Υποβολή της Αίτησης	21-04-2023:11.34	Υπάλληλος 1
B	Υποβολή της Αίτησης	21-04-2023:12.34	Υπάλληλος 1
A	Αποδοχή της Αίτησης	21-04-2023:12.40	Υπάλληλος 2
B	Αποδοχή της Αίτησης	21-04-2023:13.34	Υπάλληλος 3
Γ	Υποβολή της Αίτησης	21-04-2023:14.34	Υπάλληλος 3
B	Εύρεση του Αρχείου	21-04-2023:14.55	Υπάλληλος 2
Δ	Υποβολή της Αίτησης	22-04-2023:11.34	Υπάλληλος 1
Γ	Απόρριψη της Αίτησης	22-04-2023:13.00	Υπάλληλος 2
Γ	Ενημέρωση του Αιτούντος	22-04-2023:13.14	Υπάλληλος 3
A	Εύρεση του Αρχείου	22-04-2023:14.04	Υπάλληλος 1
Δ	Αποδοχή της Αίτησης	30-04-2023:12.05	Υπάλληλος 1
A	Ενημέρωση του Αιτούντος	30-04-2023:12.05	Υπάλληλος 2
Δ	Μη Εύρεση του Αρχείου	30-04-2023:12.09	Υπάλληλος 3
B	Ενημέρωση του Αιτούντος	30-04-2023:13.14	Υπάλληλος 1
A	Ενημέρωση του Αιτούντος	30-04-2023:13.20	Υπάλληλος 2
Δ	Ενημέρωση του Αιτούντος	30-04-2023:14.30	Υπάλληλος 1

Πίνακας 2.2: Αρχείο Καταγραφής Γεγονότων της Εταιρείας

Στον Πίνακα 2.2 φαίνεται το Αρχείο Καταγραφής Γεγονότων για τις εργασίες της εταιρείας. Κάθε γραμμή του πίνακα είναι ένα Γεγονός, ενώ κάθε πελάτης λογίζεται ως μία ξεχωριστή Υπόθεση. Βλέπουμε ότι πέρα από τα 3 χαρακτηριστικά που πρέπει να έχει υποχρεωτικά κάθε Γεγονός (Αναγνωριστικό Υπόθεσης, Δραστηριότητα, Χρονική Σήμανση), τα Γεγονότα του εν λόγω Αρχείου Καταγραφής περιέχουν επίσης το χαρακτηριστικό που δηλώνει τον Υπάλληλο της εταιρείας που ασχολήθηκε με την κάθε Δραστηριότητα. Από το Αρχείο Καταγραφής Γεγονότων προκύπτει ότι στο χρονικό διάστημα που εξετάζουμε η εταιρεία είχε 4 πελάτες, τους A, B, Γ και Δ, οι οποίοι ορίζουν 4 Υποθέσεις και ότι συνολικά συνέβησαν 16 Γεγονότα που σχετίζονται με τις Δραστηριότητες Υποβολή Αίτησης, Αποδοχή Αίτησης, Απόρριψη Αίτησης, Εύρεση Αρχείου, Μη Εύρεση Αρχείου Ενημέρωση Αιτούντος.

Το Αρχείο Καταγραφής Γεγονότων ως Πολυσύνολο Ιχνών παριστάνεται ως εξής:
 $L = \{ \langle \text{Υποβολή Αίτησης, Αποδοχή της Αίτησης, Εύρεση του Αρχείου, Ενημέρωση του Αιτούντος} \rangle^2, \langle \text{Υποβολή Αίτησης, Απόρριψη της Αίτησης, Ενημέρωση του Αιτούντος} \rangle, \langle \text{Υποβολή Αίτησης, Αποδοχή της Αίτησης, Μη Εύρεση του Αρχείου, Ενημέρωση του Αιτούντος} \rangle \}$.
 Παρατηρούμε ότι υπάρχουν δύο διαφορετικές Υποθέσεις (Πελάτης A και Πελάτης B) με το ίδιο Ίχνος, ενώ ο αριθμός Γεγονότων δεν είναι ίδιος σε όλες τις Υποθέσεις.

Σημείωση: Τα παραπάνω παραδείγματα είναι δύο πολύ απλουστευμένες περιπτώσεις που σκοπό έχουν να γίνουν καλύτερα κατανοητές οι έννοιες που ορίστηκαν στην αρχή της ενότητας. Στην πραγματικότητα τα Αρχεία Καταγραφής Δεδομένων είναι πολύ μεγαλύτερα σε όγκο δεδομένων και αφορούν πολύ πιο περίπλοκες Διεργασίες.

Έχοντας οπτικοποιήσει την έννοια του Αρχείου Καταγραφής Γεγονότων και έχοντας μάθει να αντλούμε κάποιες πρώτες πληροφορίες από αυτό, μπορούμε πλέον να περάσουμε σε λίγο πιο σύνθετες έννοιες, όπως αυτή της ποιότητας των Δεδομένων. Καθώς το Αρχείο Καταγραφής Γεγονότων αποτελεί το εναρκτήριο βήμα για την Εξόρυξη Διεργασιών και την βασική πηγή από την οποία αντλούμε γνώση για το Μοντέλο που πρόκειται να παραχθεί, είναι ιδιαίτερα σημαντικό τα Δεδομένα που αυτό περιέχει να είναι όσο το δυνατόν καλύτερα και πιο ποιοτικά. Η έννοια "ποιότητα δεδομένων" είναι αφηρημένη και δεν είναι εύκολο να την αντιληφθούμε με την φυσική μας διαίσθηση. Η ποιότητα δεδομένων είναι μία έννοια που αναπτύχθηκε για να δηλώσει ποια Δεδομένα προσφέρονται για καλύτερες αναλύσεις, δηλαδή για αναλύσεις από τις οποίες εξάγονται όσο το δυνατόν πιο χρήσιμα και ακριβή συμπεράσματα. Τα κριτήρια με τα οποία κρίνεται η ποιότητα ενός Αρχείου Καταγραφής προέκυψαν με βάση ζητήματα που παρατηρούνται συχνά και δημιουργούν προβλήματα στις αναλύσεις. Τα τέσσερα κυριότερα ζητήματα ποιότητας των δεδομένων που εμφανίζονται σε Αρχεία Καταγραφής Γεγονότων είναι η ύπαρξη ελλιπών Γεγονότων, λανθασμένων Γεγονότων, ανακριβών Γεγονότων και μη σχετικών Γεγονότων[10]. Σύμφωνα με τον van der Aalst προτείνεται η αξιολόγηση για την ποιότητα των δεδομένων σε μία κλίμακα από το 1 ως το 5 όπως αυτή φαίνεται στον Πίνακα 2.3.[2]

Βαθμολογία	Χαρακτηρισμός
5	Το Αρχείο Καταγραφής Γεγονότων είναι ύψιστης ποιότητας, δηλαδή αξιόπιστο, πλήρες και τα Γεγονότα είναι καλά ορισμένα. Τα Γεγονότα καταγράφονται με έναν αυτοματοποιημένο, συστηματικό, αξιόπιστο και ασφαλή τρόπο. Το απόρητο και η ασφάλεια των δεδομένων λαμβάνεται υπόψιν επαρκώς. Τα γεγονότα έχουν σαφή σημασιολογία. Αυτό συνεπάγεται την ύπαρξη μιας ή περισσότερων οντολογιών. Τα Γεγονότα και τα χαρακτηριστικά τους παραπέμπουν σε αυτή την οντολογία.
4	Τα Γεγονότα καταγράφονται αυτόματα και με συστηματικό και αξιόπιστο τρόπο, δηλαδή τα Αρχεία Καταγραφής είναι αξιόπιστα και πλήρη. Σε αντίθεση με τα συστήματα που λειτουργούν στο επίπεδο 3, έννοιες όπως η Υπόθεση και η Δραστηριότητα υποστηρίζονται με σαφή τρόπο.
3	Τα Γεγονότα καταγράφονται αυτόματα, αλλά δεν ακολουθείται συστηματική προσέγγιση για την καταγραφή τους. Ωστόσο, σε αντίθεση με τα αρχεία καταγραφής στο επίπεδο 2, υπάρχει κάποιο επίπεδο εγγύησης ότι τα καταγεγραμμένα γεγονότα ανταποκρίνονται στην πραγματικότητα (δηλαδή, το αρχείο είναι αξιόπιστο, αλλά όχι απαραίτητα πλήρες).
2	Τα Γεγονότα καταγράφονται αυτόματα, δηλαδή ως υποπροϊόν κάποιου συστήματος πληροφοριών. Η κάλυψη ποικίλλει, δηλαδή δεν ακολουθείται συστηματική προσέγγιση για να αποφασιστεί ποια γεγονότα καταγράφονται. Επιπλέον, είναι δυνατή η παράκαμψη το σύστημα πληροφοριών. Ως εκ τούτου, μπορεί να λείπουν ή να μην καταγράφονται γεγονότα σωστά.
1	Τα Αρχεία Καταγραφής Γεγονότων είναι κακής ποιότητας. Τα καταγεγραμμένα γεγονότα μπορεί να μην ανταποκρίνονται στην πραγματικότητα και μπορεί να λείπουν γεγονότα. Τέτοια χαρακτηριστικά έχουν συνήθως Αρχεία Καταγραφής Γεγονότων τα οποία καταγράφονται με το χέρι. Παραδείγματα: ίχνη που αφήνονται σε έντυπα έγγραφα που διακινούνται μέσω του οργανισμού ("κίτρινες σημειώσεις"), ιατρικά αρχεία σε χαρτί κ.λπ.

Πίνακας 2.3: Αξιολόγηση της ποιότητας των Αρχείων Καταγραφής Γεγονότων κατά τον van der Aalst

Στο υπόλοιπο μέρος της παρούσας εργασίας δεν θα ασχοληθούμε με την αξιολόγηση των Δεδομένων προς ανάλυση και γίνεται η παραδοχή ότι το Αρχείο Καταγραφής Γεγονότων περιέχει δεδομένα επιπέδου 5 στην κλίμακα του Van der Aalst.

Όπως έχει αναφερθεί σκοπός της Εξόρυξης Διεργασιών είναι η παραγωγή ενός Μοντέλου που περιγράφει την Διεργασία στα πλαίσια της οποίας έχουν συμβεί τα Γεγονότα που περιέχονται στο Αρχείο Καταγραφής. Το Μοντέλο της Διεργασίας παριστάνεται ως ένα διάγραμμα ροής, το οποίο δείχνει τα βήματα της Διεργασίας και την μετάβαση από το ένα βήμα στο άλλο. Για την αναπαράσταση του Μοντέλου χρησιμοποιούνται κυρίως τα γραφήματα άμεσης ακολουθίας (Directly-follows graphs), τα οποία είναι εύκολο να κατανοηθούν διαισθητικά και τα Δίκτυα Petri, τα οποία δύνανται να περιγράψουν πληρέστερα ιδιότητες του Μοντέλου.

Ορισμός 2.5 (Γράφημα Άμεσης Ακολουθίας). Ένα Γράφημα Άμεσης Ακολουθίας είναι ένα

ζεύγος $G = (A, F)$, όπου $A \subseteq U_{act}$ είναι ένα σύνολο Δραστηριοτήτων και $F \in \mathcal{B}((A \times A) \cup (\blacktriangleright \times A) \cup (A \times \blacksquare) \cup (\blacktriangleright \times \blacksquare))$ είναι ένα πολυσύνολο τόξων.⁴ Ο κόμβος με το χαρακτηριστικό \blacktriangleright είναι ο κόμβος έναρξης και ο κόμβος με το χαρακτηριστικό \blacksquare είναι ο κόμβος λήξης. Ισχύει επίσης $(\blacktriangleright, \blacksquare) \cap U_{act} = \emptyset$. Το σύνολο των Γραφημάτων Άμεσης Ακολουθίας U_G είναι υποσύνολο του συνόλου των μοντέλων της διεργασίας U_M . [8]

Ο Ορισμός 2.5 μπορεί να γίνει πιο κατανοητός με την βοήθεια της Εικόνας 2.2, στην οποία απεικονίζεται ένα Γράφημα Άμεσης Ακολουθίας. Οι Δραστηριότητες a, b, c, d, e του γραφήματος παριστάνονται με κόμβους στους οποίους αναγράφεται το όνομα της κάθε Δραστηριότητας και συνδέονται μεταξύ τους με τόξα. Στο Γράφημα υπάρχουν επίσης οι κόμβοι με τα σύμβολα \blacktriangleright και \blacksquare , οι οποίοι σημαίνουν αντίστοιχα την έναρξη και την λήξη της Διεργασίας και συνδέονται αντίστοιχα με τις Δραστηριότητες a και b, ενώ οι ίδιοι δεν αναπαριστούν κάποια Δραστηριότητα (δηλαδή δεν ανήκουν στο U_{act}).

Ορισμός 2.6 (Δίκτυο Petri). Έστω P ένα Σύνολο Θέσεων (παριστάνονται με κύκλους), T ένα Σύνολο Μεταβάσεων (παριστάνονται με τετράγωνα) και $F \subseteq (P \times T) \cup (T \times P)$ είναι μια Σχέση Ροής μεταξύ Θέσεων και Μεταβάσεων (και μεταξύ Μεταβάσεων και Θέσεων). Ένα Δίκτυο Petri N είναι μια Πλειάδα $N = (P, T, F)$. [11]

Ορισμός 2.7 (Επισημασμένο Δίκτυο Petri). Ένα επισημασμένο Δίκτυο Petri (Labeled Petri net) είναι μία πλειάδα (P, T, F, l) , όπου P το Σύνολο Θέσεων, T το Σύνολο Μεταβάσεων, $P \cap T \neq \emptyset$, $F \subseteq (P \times T) \cup (T \times P)$ η Σχέση Ροής και $l \in T \rightarrow U_{act}$ ⁵ η συνάρτηση επισήμανσης. Συμβολίζουμε $l(t) = \tau$ αν $t \in T \setminus \text{dom}(l)$.⁶ [8]

Ορισμός 2.8 (Αποδεκτό Δίκτυο Petri). Ένα αποδεκτό Δίκτυο Petri (Accepting Petri net) είναι μία τριάδα $AN = (N, M_{init}, M_{final})$ όπου $N = (P, T, F, l)$ είναι ένα επισημασμένο Δίκτυο Petri, $M_{init} \in \mathcal{B}(P)$ είναι η αρχική επισήμανση και $M_{final} \in \mathcal{B}(P)$ η τελική επισήμανση. Το $U_{AN} \subseteq U_M$ είναι το σύνολο των Αποδεκτών Δικτύων Petri. [8]

Στην Εικόνα 2.1 φαίνεται το Αποδεκτό Δίκτυο Petri $N = (P, T, F, l)$ όπου $P = \{p1, p2, p3, p4, p5\}$, (5 θέσεις), $T = \{t1, t2, t3, t4\}$ (4 Μεταθέσεις), $F = \{(p1, t1), (p1, t3), (p2, t3), (p2, t2), (t1, p4), (t3, p4), (t3, p3), (t2, p3), (p3, t4), (t4, p5)\}$ (9 βέλη) και $l = \{(t1, a), (t2, b), (t3, c), (t4, d)\}$ (Συνάρτηση επισήμανσης).

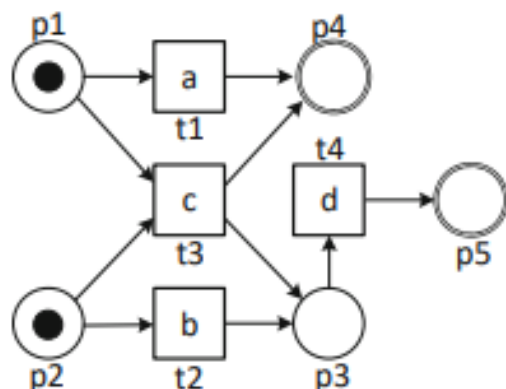
Ξεκινάμε από την αρχική σήμανση, η οποία συμβολίζεται με $\{\bullet\}$, και προχωράμε από μία σήμανση στην επόμενη πυροδοτώντας ενεργοποιημένες μεταβάσεις. Αρχικά είναι ενεργοποιημένες οι μεταβάσεις $t1, t2, t3$, διότι οι $p1$ και $p2$ (που είναι οι αρχικές σημάνσεις) συνδέονται με αυτές μέσω βελών. Η πυροδότηση του $t1$ (δηλαδή η πραγματοποίηση της Δραστηριότητας a) συνεπάγεται ότι πλέον οι $p2, p4$ θεωρούνται επισημασμένες⁷, ενώ οι

⁴Τα χαρακτηριστικά \blacktriangleright και \blacksquare δεν είναι στοιχεία του συνόλου A , αλλά προστίθενται για να δηλωθεί η έναρξη και η λήξη μιας διεργασίας. Οι 4 δυνατοί τύποι τόξου είναι οι $(a_1, a_2), (\blacktriangleright, a), (a, \blacksquare), (\blacktriangleright, \blacksquare)$, όπου $a, a_1, a_2 \in A$

⁵Το σύμβολο \rightarrow υποδηλώνει ότι η l είναι μερική συνάρτηση, δηλαδή υπάρχουν στοιχεία του T τα οποία δεν αντιστοιχίζονται μέσω της l σε στοιχεία του U_{act} .

⁶Με $\text{dom}(l)$ συμβολίζουμε το πεδίο ορισμού της l , δηλαδή τα στοιχεία t για τα οποία υπάρχει $\tau \in U_{act}$ τέτοιο ώστε $l(t) = \tau$.

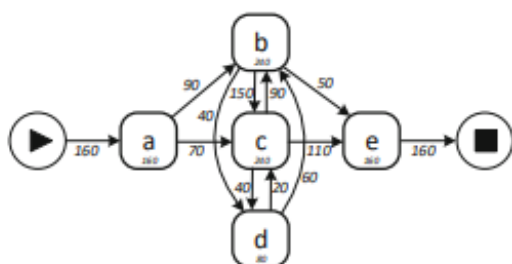
⁷Όταν αναφέρεται ο όρος επισημασμένη Μετάβαση, τότε εννοούμε ότι η εν λόγω μετάβαση φέρει κάποια ετικέτα στο Δίκτυο Petri, κάτι που δεν αλλάζει καθώς διαπερνά κανείς το Γράφημα (π.χ. η Μετάβαση $t1$ είναι επισημασμένη, καθώς φέρει την ετικέτα "a"). Όταν αναφέρεται ο όρος επισημασμένη θέση, εννοείται ότι καθώς διαπερνά κανείς το Δίκτυο Petri μαρκάρει την εν λόγω θέση. Για την καλύτερη κατανόηση βοηθούν οι αγγλικοί όροι Labeled Transition και marked Place



Εικόνα 2.1: Αποδεκτό Δίκτυο Petri

$t1$ και $t3$ παύουν να είναι ενεργοποιημένες και η μόνη ενεργοποιημένη μετάβαση είναι η $t2$. Αντίστοιχα, η πυροδότηση του $t2$ συνεπάγεται την επισήμανση των $p1, p3$ και ενεργοποιημένη μένει μόνο η $t1$ και η πυροδότηση του $t3$ συνεπάγεται την επισήμανση των $p3, p4$ με ενεργοποιημένη μετάβαση την $t4$.

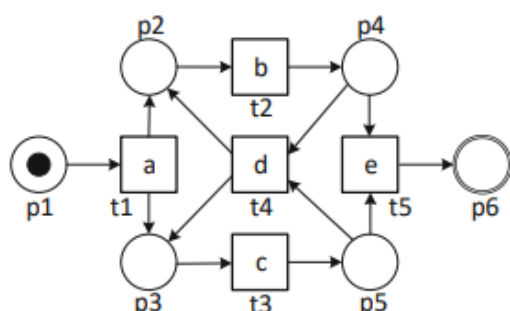
Έστω ότι πυροδοτήθηκε η $t1$. Τότε όπως αναφέρθηκε νωρίτερα ενεργοποιημένη παραμένει μόνο η $t2$. Αν αυτή πυροδοτηθεί, τότε επισημαίνονται οι $p3, p4$ (η $p4$ ήταν ήδη επισημασμένη από το προηγούμενο βήμα όπου πυροδοτήθηκε η $t1$). Με αυτή την επισήμανση μένει ως μόνη ενεργοποιημένη μετάβαση η $t4$, η πυροδότηση της οποίας συνεπάγεται την επισήμανση των $p4, p5$, που αποτελεί τελική επισήμανση. Η ακολουθία των επισημασμένων μεταβάσεων που προκύπτει από την μία διαπέραση του Δικτύου με αφετηρία μία αρχική επισήμανση και πέρας μία τελική επισήμανση αποτελεί Ίχνος του αντίστοιχου Αρχείου Καταγραφής Γεγονότων. [8]



Εικόνα 2.2: Γράφημα Άμεσης Ακολουθίας για το Αρχείο Καταγραφής Γεγονότων L_1

Το Γράφημα Άμεσης Ακολουθίας της Εικόνας 2.2 και το Αποδεκτό Δίκτυο Petri της Εικόνας 2.3 παριστάνουν και τα δύο το Μοντέλο που παράγεται από το Αρχείο Καταγραφής Γεγονότων $L_1 = \{ \langle a, b, c, e \rangle^{50}, \langle a, c, b, e \rangle^{40}, \langle a, b, c, d, b, c, e \rangle^{30}, \langle a, c, b, d, b, c, e \rangle^{20}, \langle a, b, c, d, c, b, e \rangle^{10}, \langle a, c, b, d, c, b, d, b, c, e \rangle^{10} \}$ [8].

Έχοντας υπόψιν όλα τα παραπάνω μπορούμε πια να δούμε τον τρόπο με τον οποίο παράγεται το Μοντέλο μέσα από το Αρχείο Καταγραφής Γεγονότων. Για τον σκοπό αυτό χρησιμοποιούνται οι Αλγόριθμοι Ανακάλυψης του Μοντέλου. Η κεντρική ιδέα όλων των Αλγόριθμων Ανακάλυψης είναι ότι μέσα από ένα σύνολο Γεγονότων πρέπει να ανιχνεύσουν τις μεταξύ τους σχέσεις, δηλαδή την σειρά με την οποία αυτά εκτελούνται. Ωστόσο ανάλογα

Εικόνα 2.3: Δίκτυο Petri για το Αρχείο Καταγραφής Γεγονότων L_1

με τον σκοπό για τον οποίο λαμβάνει χώρα η Εξόρυξη Διεργασιών ο τρόπος λειτουργίας του Αλγόριθμου και το παραγόμενο Μοντέλο μπορεί να διαφέρουν. Για παράδειγμα μπορεί ένας αλγόριθμος να απεικονίζει κάθε συμπεριφορά που παρατηρείται στο Αρχείο Καταγραφής Γεγονότων, ακόμα κι αν η συχνότητα αυτής είναι πολύ μικρή, ενώ κάποιος άλλος αλγόριθμος μπορεί να παράγει ένα μοντέλο που δεν περιέχει συμπεριφορές που αποκλίνουν από την συμπεριφορά της πλειοψηφίας. Στην παρούσα εργασία ασχολούμαστε με δεδομένα που προέρχονται από τον τραπεζικό τομέα και για αυτό το λόγο θα περιορίσουμε την έρευνά μας σε αλγόριθμους που επεξεργάζονται επιχειρησιακές διεργασίες. Αυτοί οι αλγόριθμοι διακρίνονται κυρίως σε δύο κατηγορίες, τους «τοπικούς» και τους «σφαιρικούς». Οι τοπικοί αλγόριθμοι παράγουν μοντέλα εξετάζοντας κάθε φορά τις σχέσεις μεταξύ των δραστηριοτήτων, ενώ οι σφαιρικοί αλγόριθμοι ξεκινούν με ένα πλήρες μοντέλο, δηλαδή ένα μοντέλο που περιέχει κάθε συμπεριφορά που υπάρχει στο Αρχείο καταγραφής και σταδιακά το βελτιώνουν. [12]

Οι τρεις πιο συχνοί Αλγόριθμοι Ανακάλυψης είναι οι εξής:

1. Ευρετικός Αλγόριθμος Εξόρυξης (Heuristic miner)
2. Γενετικός Αλγόριθμος Εξόρυξης (Genetic miner)
3. Ασαφής Αλγόριθμος Εξόρυξης (Fuzzy miner)

[13] Σημαντικός είναι επίσης ο Αλγόριθμος Alpha, ο οποίος είναι παρόμοιος με τον Ευρετικό Αλγόριθμο. Στην πραγματικότητα ο δεύτερος αποτελεί επέκταση του πρώτου.

2.1.1 Ευρετικός Αλγόριθμος Εξόρυξης (Heuristic Miner)

Ο Ευρετικός Αλγόριθμος Εξόρυξης μπορεί να ανταπεξέλθει στην περίπτωση Διεργασιών με «Θόρυβο». [13] Θόρυβος ονομάζεται το σύνολο των Ιχνών που εμφανίζονται με χαμηλή συχνότητα στο Αρχείο Καταγραφής Γεγονότων και αποκλίνουν από την συνήθη συμπεριφορά του μοντέλου [2]. Συνεπώς ο Ευρετικός Αλγόριθμος μπορεί να απεικονίσει την κύρια συμπεριφορά της Διεργασίας χωρίς όλες τις λεπτομέρειες.

Τα βήματα του Αλγόριθμου είναι τα εξής:

1. Διάβασμα του Αρχείου Καταγραφής Γεγονότων

2. Εύρεση του Συνόλου των Δραστηριοτήτων (Tasks)
3. Εξαγωγή των Σχέσεων Διάταξης (Ordering Relations) βάσει των Συχνότητων τους
4. Σχεδίαση του Διαγράμματος Ροής βάσει των Σχέσεων Διάταξης που εξήχθησαν
5. Επιστροφή του Διαγράμματος Ροής που σχεδιάστηκε

Οι Σχέσεις Διάταξης αφορούν στην σειρά με την οποία εκτελούνται δύο δραστηριότητες στην Διεργασία, δηλαδή αν για δύο Δραστηριότητες a και b μπορούμε να συμπεράνουμε ότι η a έπεται της b ή το αντίστροφο ή τίποτα από τα δύο. Όπως παρατηρούμε ο αλγόριθμος αυτός εξετάζει σχέσεις μεταξύ δραστηριοτήτων, δηλαδή είναι ένας τοπικός αλγόριθμος.

Ορισμός 2.9 (Σχέσεις Διάταξης). Έστω T ένα Σύνολο Δραστηριοτήτων και $s \in T$ ένα Ίχνος Γεγονότων. Το $W \subseteq T$ είναι ένα Αρχείο Καταγραφής Γεγονότων, δηλαδή ένα πομπύσυνολο από Ίχνη Γεγονότων. Τέλος έστω $a, b \in T$. Παρακάτω ορίζονται οι σχέσεις διάταξης $a >_W b$, $a \rightarrow_W b$, $a \#_W b$, $a \parallel_W b$, $a >>_W b$, $a >>>_W b$ [14]:

- $a >_W b$ αν και μόνο αν υπάρχει Ίχνος $S = t_1, t_2, t_3, \dots, t_n$ και $i \in \{1, \dots, n - 1\}$ τέτοιο ώστε $S \in W$ και $t_i = a, t_{i+1} = b$
- $a \rightarrow_W b$ αν και μόνο αν $a >_W b$ και $b \not>_W a$
- $a \#_W b$ αν και μόνο αν $a \not>_W b$ και $b \not>_W a$
- $a \parallel_W b$ αν και μόνο αν $a >_W b$ και $b >_W a$
- $a >>_W b$ αν και μόνο αν υπάρχει Ίχνος $S = t_1, t_2, t_3, \dots, t_n$ και $i \in \{1, \dots, n - 1\}$ τέτοια ώστε $S \in W$ και $t_i = a, t_{i+1} = b, t_{i+2} = a$
- $a >>>_W b$ αν και μόνο αν υπάρχει Ίχνος $S = t_1, t_2, t_3, \dots, t_n$ και $i < j$ και $i, j \in 1, \dots, n$ τέτοια ώστε $S \in W$ και $t_i = a, t_j = b$

Στην πραγματικότητα η σχέση $a >_W b$ υποδεικνύει ότι η Δραστηριότητα b ακολουθεί άμεσα την Δραστηριότητα a , η σχέση $a \rightarrow_W b$ υποδεικνύει ότι η Δραστηριότητα b ακολουθεί άμεσα την Δραστηριότητα a αλλά δεν συμβαίνει το αντίστροφο, η σχέση $a \#_W b$ υποδεικνύει ότι οι Δραστηριότητες a και b δεν ακολουθούν ποτέ άμεσα η μία την άλλη και η σχέση $a \parallel_W b$ υποδεικνύει ότι κάθε μία από τις Δραστηριότητες a και b μπορεί να έπεται της άλλης, γεγονός που σημαίνει ότι πιθανόν αυτές οι δύο Δραστηριότητες είναι παράλληλες.

Για να παραχθεί το Διάγραμμα Ροής του Μοντέλου πρέπει να ανιχνευθεί ποια από τις παραπάνω σχέσεις ισχύει για κάθε δυάδα Δραστηριοτήτων της Διεργασίας. Όπως αναφέρθηκε στην αρχή της ενότητας στο Αρχείο Καταγραφής Γεγονότων μπορεί να υπάρχει θόρυβος, δηλαδή μπορεί να υπάρχουν Ίχνη που αφορούν Υποθέσεις οι οποίες δεν ανταποκρίνονται στον γενικό κανόνα του Μοντέλου. Συνεπώς ενδέχεται να υπάρχουν Γεγονότα που αφορούν την ίδια δυάδα Δραστηριοτήτων και σε κάποια Υπόθεση το ένα Γεγονός ακολουθεί άμεσα το άλλο, σε κάποια άλλη και τα δύο Γεγονότα ακολουθούν το ένα το άλλο, σε κάποια τρίτη συμβαίνει μόνο το ένα Γεγονός κ.ο.κ. Σκοπός του Ευρετικού Αλγόριθμου δεν είναι να αποτυπωθούν

όλες αυτές οι περιπτώσεις στο μοντέλο, αλλά οι κυριότερες, αυτές που εκφράζουν την συνήθη συμπεριφορά της Διεργασίας, χωρίς τον θόρυβο. Έτσι είναι σημαντικό να ανιχνευθεί ποια ή ποιες σχέσεις διάταξης ισχύουν για δύο Δραστηριότητες κρίνοντας με βάση τη συνηθέστερη συμπεριφορά των Γεγονότων που αντιστοιχούν στις δραστηριότητες αυτές. Αυτό μπορεί να συμβεί μέσω μίας μετρικής.

Ορισμός 2.10 (Μετρική $a \Rightarrow_w b$). Η μετρική $a \Rightarrow_w b$ ορίζεται ως ο λόγος της διαφοράς του πλήθους εμφανίσεων στο Αρχείο Καταγραφής Γεγονότων που αντιστοιχούν στην Δραστηριότητα b αμέσως μετά από γεγονότα που αντιστοιχούν στην Δραστηριότητα a από το πλήθος των Ιχνών του Αρχείου Καταγραφής Γεγονότων στα οποία ισχύει το αντίστροφο προς τον συνολικό πλήθος των Ιχνών στα οποία είτε γεγονότα που αντιστοιχούν στην Δραστηριότητα a ακολουθούν γεγονότα που αντιστοιχούν στην δραστηριότητα b είτε συμβαίνει το αντίστροφο, αυξημένο κατά 1:

$$a \Rightarrow_w b \equiv \left(\frac{|a >_w b| - |b >_w a|}{|a >_w b| + |b >_w a| + 1} \right)$$

Από τον ορισμό προκύπτει ότι η τιμή της μετρικής $a \Rightarrow_w b$ παίρνει τιμές στο διάστημα $(-1, 1)$. Όταν η τιμή της μετρικής $a \Rightarrow_w b$ είναι κοντά στο 0, σημαίνει ότι η διαφορά του πλήθους των Ιχνών στα οποία η δραστηριότητα b ακολουθεί άμεσα την δραστηριότητα a από το πλήθος των Ιχνών στα οποία ισχύει το αντίστροφο είναι πολύ μικρή σε σχέση με το συνολικό πλήθος Ιχνών στα οποία συμβαίνει κάτι από τα δύο, δηλαδή οι φορές που η δραστηριότητα a ακολουθεί άμεσα την δραστηριότητα b δεν είναι αισθητά περισσότερες από τις φορές που η δραστηριότητα b ακολουθεί την δραστηριότητα a . Όταν η μετρική $a \Rightarrow_w b$ βρίσκεται κοντά στο 1, τότε το πλήθος των Ιχνών στα οποία η δραστηριότητα a ακολουθεί άμεσα την δραστηριότητα b είναι πολύ μικρό σχετικά με το πλήθος των Ιχνών στα οποία συμβαίνει το αντίστροφο. Στις περιπτώσεις αυτές, αν η τιμή της μετρικής ξεπερνά κάποιο κατώφλι που έχουμε ορίσει (και με βάση το οποίο μία τιμή θεωρείται "αρκετά κοντά" στο 1) μπορούμε να εξαγάγουμε τον κανόνα ότι η δραστηριότητα b ακολουθεί άμεσα την δραστηριότητα a στο μοντέλο που παράγεται. Ακολουθώντας αντίστοιχο συλλογισμό, όταν η τιμή της μετρικής είναι κοντά στο -1 (δηλαδή είναι αρνητική και ξεπερνά σε απόλυτη τιμή την τιμή κατωφλίου που έχει οριστεί), τότε μπορεί να εξαχθεί το συμπέρασμα ότι στο μοντέλο που παράγεται η δραστηριότητα b ακολουθεί άμεσα την δραστηριότητα a . Είναι σαφές ότι η μετρική δεν μπορεί να πάρει τις τιμές -1 και 1.

Πέρα από τις άμεσες ακολουθίες δραστηριοτήτων το μοντέλο μπορεί να περιέχει βρόγχους επανάληψης. Για μικρούς βρόγχους μήκους 1 και 2, δηλαδή για περιπτώσεις όπου μία δραστηριότητα επαναλαμβάνεται αμέσως μετά την εκτέλεση της και για περιπτώσεις όπου μία δυάδα δραστηριοτήτων επαναλαμβάνεται μετά την εκτέλεση της δεύτερης δραστηριότητας, ισχύουν οι παρακάτω σχέσεις:

$$a \Rightarrow_w a \equiv \left(\frac{|a >_w a|}{|a >_w a| + 1} \right), \text{ για βρόγχους μήκους 1 και}$$

$$a \Rightarrow_{2w} b \equiv \left(\frac{|a >>_w b| + |b >>_w a|}{|a >>_w b| + |b >>_w a| + 1} \right), \text{ για βρόγχους μήκους 2}$$

Η μετρική $a \Rightarrow_w a$ είναι ο λόγος του πλήθους των εμφανίσεων γεγονότων που αντιστοιχούν στην δραστηριότητα a αμέσως μετά από γεγονότα του ίδιου Ίχνους που αντιστοιχούν στην ίδια δραστηριότητα στο Αρχείο Καταγραφής Γεγονότων. Η τιμή της μετρικής βρίσκεται

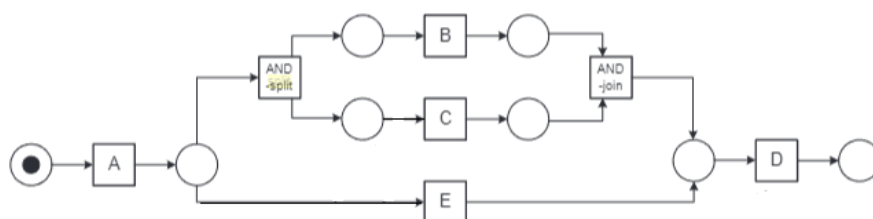
στο διάστημα $(0, 1)$. Όσο πιο κοντά είναι η τιμή της μετρικής στο 1, τόσο περισσότερες φορές εμφανίζεται η δραστηριότητα a μετά από την ίδια στο Αρχείο Καταγραφής Γεγονότων. Αν η τιμή της μετρικής ξεπερνά την τιμή κατωφλίου που έχουμε ορίσει, τότε μπορεί να εξαχθεί ο κανόνας ότι η δραστηριότητα a είναι επαναληπτική στο παραγόμενο μοντέλο.

Η μετρική $a \Rightarrow_{2W} b$ είναι ο λόγος του πλήθους εμφανίσεων γεγονότων που αντιστοιχούν σε κάποια από τις δραστηριότητες a ή b και ακολουθούν άμεσα γεγονός που αντιστοιχεί στην άλλη δραστηριότητα προς το ίδιο πλήθος αυξημένο κατά 1. Η έκφραση $|a \gg_W b| + |b \gg_W a|$ ισούται με το πλήθος των εμφανίσεων κάποιας από τις δραστηριότητες a, b αμέσως μετά την άλλη στο Αρχείο Καταγραφής Γεγονότων. Η μετρική αυτή παίρνει τιμές στο διάστημα $(0, 1)$. Όσο πιο κοντά είναι αυτός ο λόγος στο 1, τόσες περισσότερες είναι οι περιπτώσεις όπου οι a και b ακολουθούν η μία την άλλη. Αν η τιμή της μετρικής ξεπερνά την τιμή κατωφλίου που έχουμε ορίσει, τότε μπορεί να εξαχθεί ο κανόνας ότι υπάρχει επαναληπτικός βρόγχος μεταξύ των δραστηριοτήτων a και b στο παραγόμενο μοντέλο.

Παρατηρούμε ότι όταν η μετρική $a \Rightarrow_W b$ τείνει στο 1 (ή στο -1), τότε αναγκαστικά και η $a \Rightarrow_{2W} b$ τείνει στο 1, καθώς $|a \gg_W b| + |b \gg_W a| \approx |a \gg_W b|$ (ή αντίστοιχα $|a \gg_W b| + |b \gg_W a| \approx |b \gg_W a|$). Άρα ο έλεγχος της $a \Rightarrow_{2W} b$ έχει νόημα μόνο στις περιπτώσεις όπου η τιμή της $a \Rightarrow_W b$ είναι κάτω από το κατώφλι που έχουμε ορίσει και ειδικά όταν είναι κοντά στο 0.

Η τιμή κατωφλίου ορίζεται από τον αναλυτή ανάλογα με τον βαθμό στον οποίο θέλει κανείς να αποτυπώσει την Διεργασία. Συγκεκριμένα όταν η τιμή κατωφλίου μεγαλώνει, μειώνονται τα Ίχνη που περιλαμβάνονται στο μοντέλο και αποτυπώνονται οι συνηθέστερες συμπεριφορές με βάση το Αρχείο Καταγραφής Γεγονότων, ενώ όταν η τιμή κατωφλίου γίνεται μικρότερη αποτυπώνονται στο Μοντέλο περισσότερα Ίχνη. Σε κάποιες παραλλαγές του αλγόριθμου το κριτήριο δεν είναι η υπέρβαση της τιμής κατωφλίου, αλλά για κάθε δραστηριότητα a υπολογίζεται η μετρική $a \Rightarrow_W b_i$ για κάθε δραστηριότητα b_i (διαφορετική της a) που εμφανίζεται στο Αρχείο Καταγραφής Γεγονότων και ορίζεται ως άμεσα επόμενη της a η b η οποία είναι τέτοια ώστε $a \Rightarrow_W b = \max\{a \Rightarrow_W b_i, b_i \neq a\}$ [13],[15],[14]

Στην συνέχεια διερευνώνται οι τύποι των σχέσεων εξάρτησης, δηλαδή αν πρόκειται για σύζευξη (AND) ή διάζευξη (XOR).



Εικόνα 2.4: Δίκτυο Petri με σύζευξη και διάζευξη

Στην Εικόνα 2.4 φαίνεται ένα Δίκτυο Petri το οποίο αναπαριστά ένα Μοντέλο στο οποίο μετά την δραστηριότητα A ακολουθούν είτε οι δραστηριότητες B ή C (ή και οι δύο) είτε η

δραστηριότητα E. Για να δηλωθεί ότι οι δραστηριότητες B και C εκτελούνται παράλληλα (δηλαδή $A >_w B$ και $B >_w A$) προστίθενται οι μεταβάσεις AND-split και AND-join, οι οποίες δεν αντιπροσωπεύουν δραστηριότητες του Αρχείου Καταγραφής Γεγονότων (όπως συμβαίνει συνήθως με τις μεταβάσεις των Δικτύων Petri), αλλά υποδηλώνουν την σχέση σύζευξης. Τέτοιες σχέσεις ανιχνεύονται με την βοήθεια της μετρικής $A \Rightarrow_w B \wedge C$

$$A \Rightarrow_w B \wedge C = \frac{|B >_w C| + |C >_w B|}{|A >_w B| + |A >_w C| + 1}$$

Πρόκειται για τον λόγο του πλήθους εμφανίσεων στο Αρχείο Καταγραφής Γεγονότων των δραστηριοτήτων B και C ως αμέσως επόμενη δραστηριότητα η μία της άλλης προς το πλήθος εμφανίσεων στο Αρχείο Καταγραφής Γεγονότων των B και C ως αμέσως επόμενες της A αυξημένο κατά 1. Η μετρική αυτή λαμβάνει τιμές στο διάστημα (0,1). Τιμή της μετρικής $A \Rightarrow_w B \wedge C$ κοντά στο 1 σημαίνει ότι οι δραστηριότητες B και C συνδέονται με σύζευξη, ενώ τιμή κοντά στο 0 συνεπάγεται διάζευξη.

Τέλος ο Αλγόριθμος δύναται να επεκταθεί και στην ανίχνευση σχέσεων εξάρτησης, οι οποίες αφορούν Δραστηριότητες οι οποίες δεν γεινιάζουν, ωστόσο το πεδίο αυτό αποτελεί μία αρκετά ιδιαίτερη περίπτωση και η ανάλυσή του ξεφεύγει από τους σκοπούς της παρούσας διπλωματικής.[14]

2.1.2 Γενετικός Αλγόριθμος Εξόρυξης (Genetic Miner)

Ο Γενετικός Αλγόριθμος είναι μία περίπτωση «συνολικού» αλγόριθμου, καθώς έχει ως αφητηρία έναν αρχικό πληθυσμό ατόμων (δηλαδή μοντέλων διεργασιών) και σταδιακά επιλέγει τα καταλληλότερα άτομα (most fitted individuals) ή δημιουργεί νέα άτομα χρησιμοποιώντας τους γενετικούς τελεστές. [16]. Ο στόχος του Γενετικού Αλγόριθμου είναι να ανταπεξέλθει σε προβλήματα όπως η ύπαρξη Δραστηριοτήτων που εμφανίζονται πάνω από μία φορές, κρυφών Δραστηριοτήτων, περιορισμών, Θορύβου και άλλα προβλήματα που δεν μπορούν να χειριστούν μέθοδοι όπως ο Ευρετικός Αλγόριθμος.

Τα βήματα του Αλγόριθμου είναι τα εξής [13]:

1. Διάβαση του Αρχείου Καταγραφής Γεγονότων
2. Δημιουργία του Αρχικού Πληθυσμού
3. Υπολογισμός της Καταλληλότητας των Ατόμων του Αρχικού Πληθυσμού
4. Τερματισμός και Επιστροφή των Καλύτερων (Fittest) ατόμων
5. Δημιουργία του Επόμενου Πληθυσμού χρησιμοποιώντας τους Γενετικούς Τελεστές

Τα Δίκτυα Petri και τα Διαγράμματα Άμεσης Ακολουθίας που έχουν οριστεί προηγουμένως δεν είναι ιδιαίτερα βοηθητικά για την απεικόνιση του Μοντέλου στην περίπτωση του Γενετικού Αλγόριθμου, για αυτόν τον λόγο η παρουσίαση του Μοντέλου γίνεται με την χρήση

των Αιτιολογικών Πινάκων, οι οποίοι ωστόσο μπορούν να μετατραπούν σε Δίκτυα Petri τα οποία περιέχουν κρυφές μεταβάσεις, δηλαδή συζεύξεις και διαζεύξεις (βλ. Ενότητα 2.1.1).

Ορισμός 2.11 (Αιτιολογικός Πίνακας (Causal Matrix)). *Αιτιολογικός Πίνακας είναι μία Πλειάδα $CM = (A, C, I, O)$ όπου*

A είναι ένα πεπερασμένο σύνολο Δραστηριοτήτων,

$C \subseteq A \times A$ είναι μία Σχέση Αιτιότητας (Causality Relation)

$I \in A \rightarrow \mathcal{P}(\mathcal{P}(A))$ ⁸ η Συνάρτηση Συνθήκης Εισόδου

$O \in A \rightarrow \mathcal{P}(\mathcal{P}(A))$ η Συνάρτηση Συνθήκης Εξόδου

έτσι ώστε:

$$C = \{(a_1, a_2) \in A \times A \mid a_1 \in I(a_2)\}$$

$$C = \{(a_1, a_2) \in A \times A \mid a_2 \in O(a_1)\}$$

$$\forall a \in A \forall s, s' \in I(a) \ s \cap s' \neq \emptyset \Rightarrow s = s'$$

$$\forall a \in A \forall s, s' \in O(a) \ s \cap s' \neq \emptyset \Rightarrow s = s'$$

*$C \cup \{(a_o, a_i) \in A \times A \mid a_o \overset{C}{\cdot} = \emptyset \wedge \overset{C}{\cdot} a_i = \emptyset\}$ είναι ένα Ισχυρά Συνδεδεμένο Γράφημα.*⁹

Ορισμός 2.12 (Απεικόνιση Αιτιολογικού Πίνακα σε Δίκτυο Petri). *Έστω $CM = (A, C, I, O)$ ένας Αιτιολογικός Πίνακας. Η αφηρημένη (nαινε) απεικόνιση του σε Δίκτυο Petri είναι μία Πλειάδα $\Pi_{CM \rightarrow PN}^N(CM) = (P, T, F)$, όπου*

$$P = \{i, o\} \cup \{i_{t,s} \mid t \in A \wedge s \in I(t)\} \cup \{o_{t,s} \mid t \in A \wedge s \in O(t)\}$$

$$T = \{i, o\} \cup \{i_{t,s} \mid t \in A \wedge s \in I(t)\} \cup \{o_{t,s} \mid t \in A \wedge s \in O(t)\}$$

$$F = \{(i, t) \wedge t \in A \wedge \overset{C}{\cdot} t = \emptyset\} \cup \{(t, o) \mid t \in A \wedge \overset{C}{\cdot} = \emptyset\} \cup \{(i_{t,s}, t) \mid t \in A \wedge s \in I(t)\} \cup \{(t, o_{t,s}) \mid t \in A \wedge s \in O(t)\} \cup \{o_{t_1,s}, m_{t_1,t_2} \mid (t_1, t_2) \in C \wedge s \in O(t_1) \wedge t_2 \in s\} \cup \{m_{t_1,t_2}, i_{t_2,s} \mid (t_1, t_2) \in C \wedge s \in I(t_2) \wedge t_1 \in s\}$$

Παράδειγμα 2.3. *Δίνεται ο Αιτιολογικός Πίνακας:*

Δραστηριότητα	Είσοδος	Έξοδος
A	{}	{{B}, {C, D}}
B	{{A}}	{{E, F}}
C	{{A}}	{{E}}
D	{{A}}	{{F}}
E	{{B}, {C}}	{{G}}
F	{{B}, {D}}	{{G}}
G	{{E}, {F}}	{}

Πίνακας 2.4: Αιτιολογικός Πίνακας

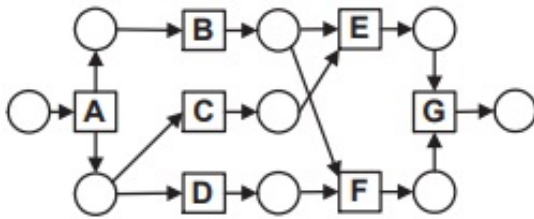
Μία απεικόνιση του Αιτιολογικού Πίνακα σε Δίκτυο Petri φαίνεται στην Εικόνα 2.5.

Η ερμηνεία του αιτιολογικού Πίνακα 2.4 και η μετατροπή του σε Δίκτυο Petri γίνεται ως εξής:

Η δραστηριότητα A έχει ως είσοδο το κενό {}, δηλαδή είναι δραστηριότητα έναρξης, και έχει ως έξοδο τις δραστηριότητες B και {C, D}, δηλαδή στο Δίκτυο Petri μετά την Δραστηριότητα A

⁸Με $\mathcal{P}(A)$ συμβολίζεται το δυναμοσύνολο του A

⁹Με $\overset{C}{\cdot} t$ συμβολίζουμε το σύνολο των Εισόδων (Inputs) που δέχεται η μετάβαση t και ανήκουν στο C, ενώ με $p \overset{C}{\cdot}$ συμβολίζουμε το σύνολο των Μεταβάσεων που δέχονται ως Είσοδο το p και ανήκουν στο C.



Εικόνα 2.5: Απεικόνιση του Αιτιολογικού Πίνακα σε Δίκτυο Petri

υπάρχει διακλάδωση 2 κλάδων (διάζευξη), όπου ο ένας κλάδος οδηγεί στην δραστηριότητα B και ο άλλος κλάδος είναι σύζευξη που οδηγεί στις C, D. Η Δραστηριότητα B δέχεται ως είσοδο την A και έχει ως έξοδο τις E, F, στις E,F οδηγούμαστε με σύζευξη. Η Δραστηριότητα C έχει ως είσοδο την A και ως έξοδο την E, δηλαδή μετά την δραστηριότητα C ακολουθεί η E. Η Δραστηριότητα D έχει ως Είσοδο την A και ως έξοδο την F, δηλαδή μετά την δραστηριότητα D ακολουθεί η F. Η Δραστηριότητα E έχει ως Είσοδο τις Δραστηριότητες B και C και ως έξοδο την G, δηλαδή μετά την Δραστηριότητα E ακολουθεί η G. Η Δραστηριότητα F έχει ως Είσοδο τις Δραστηριότητες E και F και ως έξοδο το κενό {}, δηλαδή είναι Δραστηριότητα τερματισμού. Σημείωση: Η απεικόνιση του Δικτύου Petri της Εικόνας 2.5 απεικονίζει την σύζευξη και την διάζευξη με διαφορετικό τρόπο από αυτόν που προτείνεται στην Ενότητα 2.1.1. Στην πραγματικότητα η απεικόνιση αυτή αποτελεί μία απλουστευμένη απεικόνιση Δικτύου Petri με κρυφές Μεταβάσεις. Σε περίπτωση απλών Μοντέλων χωρίς σύνθετες διακλαδώσεις, όπως η παρούσα, η απεικόνιση αυτή είναι πιο ευανάγνωστη.

Ο Γενετικός Αλγόριθμος ξεκινάει με την κατασκευή του Αρχικού Πληθυσμού, με πλήθος ίσο με τον αριθμό των Αιτιολογικών Πινάκων που μπορούν να κατασκευαστούν. Άτομα του οποίου είναι Αιτιολογικοί Πίνακες.[16] Όλα τα Άτομα του Πληθυσμού έχουν το ίδιο σύνολο Δραστηριοτήτων, το οποίο περιέχει τις Δραστηριότητες του Αρχείου Καταγραφής, αλλά οι σχέσεις αιτιότητας C και οι συναρτήσεις κατάστασης O και I δύνανται να διαφέρουν. Ο Αρχικός Πληθυσμός μπορεί να περιέχει οποιοδήποτε Άτομο στον Χώρο Αναζήτησης (Search Space), ο οποίος ορίζεται από το Σύνολο Δραστηριοτήτων. Όσο μεγαλύτερος είναι ο αριθμός Δραστηριοτήτων σε ένα Αρχείο Καταγραφής, τόσο μεγαλύτερος είναι ο Χώρος Αναζήτησης.

Η Καταλληλότητα (Fitness) ενός Ατόμου του Πληθυσμού περιγράφει τον βαθμό στον οποίο το Άτομο του Πληθυσμού περιγράφει σωστά την συμπεριφορά του Αρχείου Καταγραφής. Όσο καλύτερα μπορεί να αναλύσει ένα μοντέλο τα Ίχνη του Αρχείου Καταγραφής Γεγονότων, τόσο μεγαλύτερη είναι η τιμή της Καταλληλότητας. Συγκεκριμένα η τιμή της Καταλληλότητας δίνεται από την σχέση:

$$\text{Καταλληλότητα} = 0.40 \times \frac{\text{Αναλυμένες Δραστηριότητες}}{\text{Δραστηριότητες του Αρχείου Καταγραφής Γεγονότων}} + 0.60 \times \frac{\text{Ίχνη του Αρχείου Καταγραφής Γεγονότων που αναλύθηκαν πλήρως}}{\text{Ίχνη του Αρχείου Καταγραφής Γεγονότων}}$$

Όταν ένα Ίχνος του Αρχείου Καταγραφής Γεγονότων αναλύεται από ένα άτομο του πληθυσμού (δηλαδή έναν αιτιολογικό πίνακα) ενδεχομένως κάποια δραστηριότητα να μην μπορεί να αναλυθεί από το άτομο αυτό, δηλαδή το υπό εξέταση Ίχνος να μην καλύπτεται από το εν λόγω μοντέλο, τότε η ανάλυση του Ίχνους συνεχίζεται με τις επόμενες Δραστηριότητες του

Ίχνους. Με τον όρο Αναλυμένες Δραστηριότητες εννοούνται οι Δραστηριότητες που έχουν μπορεί να αναλυθούν από το άτομο του πληθυσμού και με τον όσο πλήρως αναλυμένα Ίχνη εννοούνται τα Ίχνη του Αρχείου Καταγραφής για τα οποία όλες οι δραστηριότητες που περιέχονται σε αυτά έχουν αναλυθεί από το άτομο του πληθυσμού. Συνεπώς η Καταλληλότητα είναι ένα μέτρο που μας δείχνει πόσα Ίχνη του Αρχείου Καταγραφής καλύπτονται από το μοντέλο που προτείνει ο εκάστοτε αιτιολογικός πίνακας.

Οι τιμές της Καταλληλότητας βρίσκονται στο διάστημα $(0, 1]$. Η τιμή 1 μπορεί να επιτευχθεί μόνο σε Αρχεία Καταγραφής Γεγονότων που δεν περιέχουν θόρυβο, ενώ σε Αρχεία Καταγραφής Γεγονότων με θόρυβο μας ενδιαφέρουν τιμές της Καταλληλότητας όσο το δυνατόν πιο κοντά στο 1. [16]

Ο Αλγόριθμος χρησιμοποιεί τους τελεστές Ελιτισμός (Elitism), Διασταύρωση (Crossover) και Μεταλλαγή (Mutation). Μέσω του Ελιτισμού ένα ποσοστό του καλύτερα προσαρμοσμένου ατόμου του πληθυσμού αντιγράφεται στην επόμενη γενιά του πληθυσμού. Η Διασταύρωση δημιουργεί νέα άτομα του πληθυσμού (απόγονοι) βάσει των καλύτερα προσαρμοσμένων ατόμων του παρόντος πληθυσμού (γονείς). Κατά την Μεταλλαγή δύναται να αλλάξουν κάποια στοιχεία του πληθυσμού.

Πιο αναλυτικά, κάθε φορά επιλέγονται τυχαία 5 άτομα του πληθυσμού, από τα οποία μέσω μίας μεθόδου "τουρνουά", όπου κερδίζει το άτομο με την μεγαλύτερη τιμή Καταλληλότητας, επιλέγονται οι δύο γονείς. Ο τελεστής Διασταύρωση επιλέγει τυχαία κάποια Δραστηριότητα ως σημείο της διασταύρωσης, ώστε στο σημείο αυτό να ανασυνδυαστούν οι εισόδοι και έξοδοι των δύο γονέων με σκοπό την δημιουργία των δύο απογόνων. Ο ανασυνδυασμός αυτός μπορεί να συνεπάγεται για ένα άτομο (για έναν γονέα) να χάνει δραστηριότητες από τα υποσύνολα στις συναρτήσεις συνθήκης εισόδου/εξόδου, να προστίθενται δραστηριότητες στα υποσύνολα στις συναρτήσεις συνθήκης εισόδου/εξόδου του, να ανταλλάσσει σχέσεις αιτιότητας με άλλα άτομα, να ενσωματώνει σχέσεις αιτιότητας που βρίσκονται στον πληθυσμό αλλά όχι στο άτομο, να χάνει σχέσεις αιτιότητας, να μειώνεται ή/και να αυξάνεται ο αριθμός των υποσυνόλων στις συναρτήσεις συνθηκών εισόδου/εξόδου του. Έτσι δημιουργούνται 2 απόγονοι με τρόπο ώστε ο συνολικός αριθμός των σχέσεων αιτιότητας να μένει σταθερός, αλλά να διαφέρει ο τρόπος με τον οποίο αυτές εμφανίζονται. Ο τελεστής Μεταλλαγή που ακολουθεί έχει ως στόχο την εισαγωγή νέου υλικού στον πληθυσμό μέσω της αλλαγής των σχέσεων αιτιότητας που υπάρχουν στον πληθυσμό. Αυτό γίνεται με την τυχαία προσθήκη ή αφαίρεση Δραστηριοτήτων σε υποσύνολα των συναρτήσεων συνθήκης εισόδου και εξόδου και μέσω της ανακατανομής των στοιχείων στα υποσύνολα των συναρτήσεων συνθήκης εισόδου και εξόδου.¹⁰ Ο Γενετικός Αλγόριθμος τερματίζει όταν βρεθεί ένα άτομο του πληθυσμού με Καταλληλότητα 1 ή όταν έχει υπολογιστεί ο μέγιστος επιτρεπτός αριθμός γενιών (όπως αυτός έχει οριστεί από τον αναλυτή) ή όταν το καλύτερα προσαρμοσμένο άτομο του πληθυσμού δεν έχει αλλάξει για $n/2$ συνεχόμενες γενιές (όπου n ο μέγιστος επιτρεπτός αριθμός γενιών). Διαφορετικά ο Αλγόριθμος δημιουργεί νέο πληθυσμό χρησιμοποιώντας την Καταλληλότητα και τους τελεστές Ελιτισμός, Διασταύρωση, Μεταλλαγή. [16], [13], [17]

¹⁰ Η διαφορά μεταξύ των τελεστών Διασταύρωση και Μεταλλαγή είναι ότι ο πρώτος στοχεύει στον ανασυνδυασμό του υπάρχοντος υλικού, ενώ ο δεύτερος στοχεύει στην δημιουργία νέου υλικού.

2.1.3 Ασαφής Αλγόριθμος Εξόρυξης (Fuzzy Miner)

Ο Ασαφής Αλγόριθμος μπορεί να παράξει απλοποιημένα μοντέλα για Διεργασίες οι οποίες δεν είναι καλά δομημένες και παρουσιάζουν έναν μεγάλο αριθμό μη δομημένων και αντικρουόμενων συμπεριφορών, όπως για παράδειγμα Διεργασίες τύπου σπαγκέτι(βλ. Εικόνα 2.6.¹¹ Αυτό επιτυγχάνεται με την χρήση τεχνικών για την αφαίρεση ασήμαντων ακμών, την ομαδοποίηση κόμβων με μεγάλη συσχέτιση κ.α. [13]



Εικόνα 2.6: Απεικόνιση μίας Διεργασίας τύπου σπαγκέτι. [1]

Ο Αλγόριθμος δέχεται ως Είσοδο ένα σύνολο N συναλλαγών (transactions) κάθε μία από τις οποίες έχει n χαρακτηριστικά (attributes), γλωσσολογικούς όρους για ποσοτικά χαρακτηριστικά, την Ελάχιστη Ασαφή Υποστήριξη (Minimum Fuzzy Support) που καθορίζεται από τον χρήστη, την Ελάχιστη Ασαφή Εμπιστοσύνη (Minimum Fuzzy Confidence) που καθορίζεται από τον χρήστη και μία Οντολογία του Τομέα (Domain Ontology) Η έξοδος του Αλγόριθμου είναι η σύζευξη χαρακτηριστικών με παρόμοια συμπεριφορά, η δημιουργία μετα-κανόνων, η δημιουργία συχνών ασαφών συνόλων στοιχείων (frequent fuzzy itemsets) και η δημιουργία ασαφών κανόνων συσχέτισης. [13] Για να επιτευχθεί η απλοποίηση σύνθετων Διεργασιών χρησιμοποιούνται οι μετρικές Σημαντικότητα (Significance) και η Συσχέτιση (Correlation). Η Σημαντικότητα μετράει την σχετική σημασία της συμπεριφοράς, δηλαδή καθορίζει το επίπεδο ενδιαφέροντος για τα Γεγονότα ή για την σειρά εμφάνισης των Γεγονότων (π.χ. Γεγονότα που παρατηρούνται συχνότερα θεωρούνται σημαντικότερα). Η Συσχέτιση μετράει πόσο στενά σχετίζονται δύο Γεγονότα που ακολουθούν το ένα το άλλο, υπολογίζοντας μεταξύ άλλων χαρακτηριστικά που σχετίζονται με τα δύο γεγονότα και είναι κοινά ή συγκρίνοντας την ομοιότητα των ονομασιών των Γεγονότων. Χρησιμοποιώντας τις μετρικές αυτές μέσω του Αλγόριθμου διατηρούνται οι συμπεριφορές με μεγάλη Σημαντικότητα, συγχωνεύονται συμπεριφορές με μικρή Σημαντικότητα και μεγάλη Συσχέτιση και παραλείπονται συμπεριφορές με μικρή Σημαντικότητα και μικρή Συσχέτιση και έτσι προκύπτει το απλοποιημένο μοντέλο. Στο σημείο αυτό δύναται να χρησιμοποιηθεί και η έννοια της Έμφασης (emphasis), ώστε να υπογραμμιστεί η πιο σημαντική συμπεριφορά.[1]

¹¹Διεργασίες τύπου σπαγκέτι ονομάζονται οι Διεργασίες, οι οποίες κατά την απεικόνισή τους σε διάγραμμα ροής έχουν μεγάλο όγκο ακμών και μονοπατιών, στις οποίες είναι δύσκολο να διακριθούν τα σημαντικά στοιχεία της Διεργασίας

Οι Αλγόριθμοι Ανακάλυψης που παρουσιάστηκαν στο παρόν κεφάλαιο είναι τρεις βασικοί Αλγόριθμοι που καθιστούν κατανοητή την διαδικασία Ανακάλυψης του Μοντέλου. Έχουν αναπτυχθεί και άλλοι Αλγόριθμοι καθώς και παραλλαγές και βελτιώσεις των παραπάνω, ωστόσο η περαιτέρω ανάπτυξη του θέματος αυτού δεν εμπίπτει στους σκοπούς της παρούσας Διπλωματικής Εργασίας.

2.2 Η Προσαρμογή του Μοντέλου

Ο Έλεγχος Προσαρμογής του Μοντέλου (Conformance Checking) εξετάζει τις αποκλίσεις μεταξύ ενός μοντέλου και της πραγματικότητας. Οι δύο βασικές κατηγορίες του Ελέγχου Προσαρμογής είναι ο Έλεγχος Προσαρμογής μεταξύ ενός μοντέλου και ενός Αρχείου Καταγραφής Γεγονότων, όπου εξετάζεται σε ποια σημεία και με ποιο τρόπο αποκλίνει η πραγματική εκτέλεση μιας Διεργασίας από τον τρόπο με τον οποίο αυτή αποτυπώνεται σε ένα μοντέλο και ο Έλεγχος Προσαρμογής μεταξύ δύο μοντέλων, όπου εξετάζεται κατά πόσο ένα μοντέλο ανταποκρίνεται σε κάποιο άλλο μοντέλο που έχει σχεδιαστεί για την ίδια Διεργασία. [18]

2.2.1 Έλεγχος Προσαρμογής μεταξύ ενός μοντέλου και ενός Αρχείου Καταγραφής Γεγονότων

Σκοπός του Ελέγχου Προσαρμογής μεταξύ ενός μοντέλου και ενός Αρχείου Καταγραφής Γεγονότων είναι να εξεταστεί σε ποιο βαθμό η πραγματικότητα, όπως αυτή καταγράφεται στο Αρχείο Καταγραφής, αποτυπώνεται επαρκώς στο μοντέλο ή αντιστρόφως σε ποιο βαθμό το μοντέλο μιας διεργασίας αποτυπώνει την διεργασία αυτή. Το αποτέλεσμα του Ελέγχου Προσαρμογής είναι αφενός η ποσοτικοποίηση του βαθμού στον οποίο το μοντέλο και η πραγματικότητα αποκλίνουν και αφετέρου ο εντοπισμός των αποκλίσεων. [2], [19]

Κατά τον Έλεγχο Προσαρμογής ποσοτικοποιούνται οι έννοιες της Προσαρμογής (Fitness) και της Ακρίβειας (Precision) του Μοντέλου χρησιμοποιώντας ως βασικούς άξονες Κανόνες Συμπεριφοράς που επιβάλλονται από το Μοντέλο, αλλά παραβιάζονται από το Αρχείο Καταγραφής, Γεγονότα που έχουν αναπαρασταθεί με λάθος τρόπο ή έχουν παραλειφθεί και μία Ευθυγράμμιση μεταξύ των Γεγονότων του Αρχείου Καταγραφής και των Δραστηριοτήτων μιας Εκτέλεσης. Υπάρχουν διαφορετικές μέθοδοι και οπτικές για την πραγματοποίηση του ελέγχου προσαρμογής. Οι κυριότερες μέθοδοι είναι η Μέθοδος των Συμβόλων και η Μέθοδος της Ευθυγράμμισης. [20], [18]

Ο Έλεγχος με την Μέθοδο των Συμβόλων θέτει ως βασική υπόθεση ότι κάθε Ίχνος του Αρχείου Καταγραφής Γεγονότων αντιστοιχεί σε μία έγκυρη ακολουθία Εκτελέσεων στο Μοντέλο. Η μέθοδος χρησιμοποιεί την απεικόνιση του Μοντέλου μέσω Συμβόλων (tokens) τα οποία παράγονται ή/και καταναλώνονται κατά την εκτέλεση Δραστηριοτήτων και αποφαινεται για την Προσαρμογή κρίνοντας από τα εναπομείναντα ή ελλιπή Σύμβολα.[20]

Η πλήρης περιγραφή της εν λόγω απεικόνισης του Μοντέλου ξεφεύγει από τους σκοπούς της παρούσας Εργασίας και για αυτό η εκτενέστερη παρουσίαση της Μεθόδου των Συμβόλων παραλείπεται.

Η Μέθοδος της Ευθυγράμμισης είναι η πιο διαδεδομένη Μέθοδος για τον Έλεγχο Προσαρμογής του Μοντέλου και βασίζεται στην Εύρεση της Βέλτιστης Ευθυγράμμισης

ανάμεσα σε κάθε Ίχνος του Αρχείου Καταγραφής και της πλησιέστερης του Συμπεριφοράς στο Μοντέλο.[7] Παρακάτω περιγράφεται αναλυτικά η Μέθοδος της Ευθυγράμμισης:

$$\gamma_1 = \begin{array}{|c|c|c|c|c|} \hline a & c & d & e & h \\ \hline a & c & d & e & h \\ \hline \end{array}, \quad \gamma_2 = \begin{array}{|c|c|c|c|c|} \hline a & d & c & e & h \\ \hline a & \gg & c & e & h \\ \hline \end{array}, \quad \gamma_3 = \begin{array}{|c|c|c|c|c|c|c|c|c|} \hline a & d & c & e & f & d & b & e & h \\ \hline a & \gg & c & e & f_2 & d & b & e & h \\ \hline \end{array}$$

Οι γ_1 , γ_2 και γ_3 είναι τρεις Ευθυγραμμίσεις μεταξύ του Μοντέλου M και του Αρχείου Καταγραφής. Στην πρώτη γραμμή κάθε Ευθυγράμμισης φαίνεται το Ίχνος του Αρχείου Καταγραφής και στην δεύτερη γραμμή φαίνεται η πλησιέστερη σε αυτό Ακολουθία Δραστηριοτήτων του Μοντέλου. Η γ_1 είναι μία τέλεια Ευθυγράμμιση για το Ίχνος $\langle a, c, d, e, h \rangle$ και το Μοντέλο M , καθώς το Ίχνος αναπαρίσταται ακριβώς στο Μοντέλο. Η γ_2 είναι η βέλτιστη Ευθυγράμμιση για το Ίχνος $\langle a, d, c, e, h \rangle$ και το Μοντέλο M . Στην δεύτερη θέση της Ευθυγράμμισης υπάρχει μία Μετατόπιση του Ίχνους, η οποία δεν μπορεί να αναπαρασταθεί από το Μοντέλο. Η Μετατόπιση αυτή συμβολίζεται με (d, \gg) . Η γ_3 είναι μία βέλτιστη Ευθυγράμμιση για το Ίχνος $\langle a, d, c, e, f, d, b, e, h \rangle$ στο Μοντέλο M . Το Γεγονός f του Αρχείου Καταγραφής ανταποκρίνεται στην Δραστηριότητα f_2 του Μοντέλου. Οι Ευθυγραμμίσεις γ_2 και γ_3 είναι περιπτώσεις μη Προσαρμογής μεταξύ του Μοντέλου και του Αρχείου Καταγραφής.[7]

Ορισμός 2.13 (Μετατοπίσεις Ευθυγράμμισης (Alignment Moves)). Έστω $N = (P, T, F)$ ένα Δίκτυο Petri (όπου P το σύνολο θέσεων, T το σύνολο μεταβάσεων και F η σχέση ροής) και L ένα Αρχείο Καταγραφής Γεγονότων. Μία επιτρεπτή Κίνηση Ευθυγράμμισης για το N και το L αναπαρίσταται ως ένα ζεύγος $(s_L, s_M) \in (T \cup \{\gg\}) \times T \cup \{\gg\} \setminus \{(\gg, \gg)\}$ τέτοιο ώστε:
 (s_L, s_M) είναι μία Μετατόπιση στο Αρχείο Καταγραφής αν $s_L \neq \gg$ και $s_M = \gg$
 (s_L, s_M) είναι μία Μετατόπιση στο Μοντέλο αν $s_L = \gg$ και $s_M \in T$
 (s_L, s_M) είναι μία Σύγχρονη Μετατόπιση.[11]

Ορισμός 2.14 (Ευθυγράμμιση (Alignment)). Έστω $N = (P, T, F)$ ένα Δίκτυο Petri με όλες τις Σημάνσεις Έναρξης (Initial Marking) και τις Σημάνσεις Λήξης (Final Markings) να συμβολίζονται ως m_i και m_f αντίστοιχα. Έστω L ένα Αρχείο Καταγραφής Γεγονότων. Έστω L_{A_N} το σύνολο όλων των Μετατοπίσεων Ευθυγράμμισης για το N και το L . Έστω $\sigma_L \in L$ ένα Ίχνος του Αρχείου Καταγραφής. Η ακολουθία $\gamma \in L_{A_N}^*$ ¹² είναι μία ευθυγράμμιση του N και του σ_L αν, αγνοώντας όλες τις εμφανίσεις του \gg , η προβολή στο πρώτο στοιχείο δίνει το σ_L και η προβολή πάνω στο δεύτερο στοιχείο δίνει μία ακολουθία $\sigma'' \in T^*$ ¹³ τέτοια ώστε $m_i \xrightarrow{\sigma''} m_f$ ¹⁴. [11]

Μετατόπιση στο Αρχείο Καταγραφής για μία μετάβαση t σημαίνει ότι η t συνέβη ενώ δεν επιτρεπόταν και μία Μετατόπιση στο Μοντέλο σημαίνει ότι η t δεν προέκυψε ενώ αναμενόταν να προκύψει.[11]. Είναι σαφές ότι κανείς μπορεί να βρει πάνω από μία διαφορετικές Ευθυγραμμίσεις μεταξύ ενός Ίχνους και ενός Μοντέλου κάθε μία από τις οποίες μπορεί να εμφανίζει διαφορετικό αριθμό αποκλίσεων. Σκοπός είναι να βρεθεί η Βέλτιστη Ευθυγράμμιση.

¹²Με $L_{A_N}^*$ συμβολίζεται το σύνολο όλων των πεπερασμένων ακολουθιών από στοιχεία του L_{A_N} .

¹³Με T^* συμβολίζεται το σύνολο όλων των πεπερασμένων ακολουθιών από στοιχεία του T .

¹⁴Ο συμβολισμός $m_i \xrightarrow{\sigma''} m_f$ υποδηλώνει ότι με αφετηρία το m_i φτάνουμε στο m_f μέσω της σ'' .

Ορισμός 2.15 (Βέλτιστη Ευθυγράμμιση). Έστω A ένα σύνολο Δραστηριοτήτων. Έστω $\sigma \in A^*$ ¹⁵ είναι ένα Ίχνος πάνω στο A και έστω $N = (P, T, F, a, m_i, m_f)$ ένα Δίκτυο Petri, όπου m_i και m_f η Σήμανση Έναρξης και Λήξης του αντίστοιχα. Έστω $lc : (A \cup \gg) \times (T \cup \gg) \rightarrow \mathbb{R}$ η συνάρτηση απόδοσης κόστους πιθανότητας για τις Μετατοπίσεις. Έστω $\Gamma_{\sigma, N}$ το σύνολο όλων των ευθυγραμμίσεων μεταξύ ενός Ίχνους σ και του Δικτύου Petri N . Ονομάζουμε την $\gamma \in \Gamma_{\sigma, N}$ Βέλτιστη Ευθυγράμμιση μεταξύ του σ και του N αν και μόνο αν για κάθε $\gamma' \in \Gamma_{\sigma, N} : \sum_{(a,t) \in \gamma} lc(a, t) \leq \sum_{(a',t') \in \gamma'} lc(a', t')$ [21]

Με λίγα λόγια βέλτιστη ευθυγράμμιση είναι η ευθυγράμμιση με το μικρότερο κόστος. Για την εύρεση της βέλτιστης ευθυγράμμισης εισάγεται η έννοια της Συνάρτησης Απόδοσης Κόστους Πιθανότητας.

Ορισμός 2.16 (Τυπική Συνάρτηση Απόδοσης Κόστους Πιθανότητας). Έστω A ένα σύνολο Δραστηριοτήτων. Έστω $N = (P, T, F, a, m_i, m_f)$ ένα Δίκτυο Petri για το A . Η Τυπική Συνάρτηση Απόδοσης Κόστους Πιθανότητας $lc : (A \cup \gg) \times (T \cup \gg) \rightarrow \mathbb{R}$ είναι η Συνάρτηση η οποία απεικονίζει όλες τις Μετατοπίσεις σε πραγματικές τιμές, έτσι ώστε για κάθε $(x, y) \in (A \cup \gg) \times (T \cup \gg) :$

$$lc(x, y) = 0 \text{ αν είτε } x \in A, y \in T \text{ και } x = a(y) \text{ ή } x = \gg, y \in T \text{ και } a(y) = \tau$$

$$lc(x, y) = \infty \text{ αν είτε } x \in A, y \in T \text{ και } x \neq a(y) \text{ ή } x = y = \gg$$

$$lc(x, y) = 1 \text{ σε κάθε άλλη περίπτωση [21]}$$

Η εύρεση της Βέλτιστης Ευθυγράμμισης είναι ένα σύνθετο πρόβλημα για την επίλυση του οποίου έχουν διαμορφωθεί διάφοροι Αλγόριθμοι και μεθοδολογίες, στην περαιτέρω ανάλυση των οποίων δεν θα προχωρήσουμε. Εφόσον βρεθεί η Βέλτιστη Ευθυγράμμιση υπάρχει και μία σαφής εικόνα για τα σημεία στα οποία αποκλίνει το Αρχείο Καταγραφής Γεγονότων από το Μοντέλο.

Με την χρήση των παραπάνω Μεθόδων γίνεται δυνατό κανείς να έχει μία σαφή εικόνα του επιπέδου στο οποίο το Μοντέλο που έχει παραχθεί περιγράφει την Πραγματικότητα και με βάση την εικόνα αυτή να προσαρμοστεί το μοντέλο.

2.2.2 Έλεγχος Προσαρμογής μεταξύ δύο μοντέλων

Σκοπός του Ελέγχου Προσαρμογής μεταξύ μοντέλων είναι να εξεταστούν οι ομοιότητες και διαφορές μεταξύ δύο μοντέλων μίας διεργασίας.

Κατά τον Έλεγχο Προσαρμογής μεταξύ Μοντέλων πρέπει αρχικά να εντοπιστούν οι κόμβοι που αναπαριστούν τις ίδιες δραστηριότητες στα δύο μοντέλα. [22] Αυτό δεν είναι πάντα μία τετριμμένη διαδικασία, καθώς τα δύο μοντέλα μπορεί να αναπαριστούν τις ίδιες δραστηριότητες με διαφορετικούς τρόπους, παραδείγματος χάριν το ένα να χρησιμοποιεί το όνομα της δραστηριότητας που συνέβη (‘Δημιουργία Αίτησης’) και το άλλο να χρησιμοποιεί τον αντίστοιχο ρηματικό τύπο (‘Δημιουργήθηκε Αίτηση’) ή εκφράσεις που στην πραγματικότητα είναι συνώνυμες (π.χ. ‘Υπογραφή συμβολαίου έναρξης διαδικασίας’ και ‘Έναρξη διαδικασίας’). Στο στάδιο αυτό ορίζεται μία συνάρτηση $map : A_0 \rightarrow A_1$, η οποία απεικονίζει δραστηριότητες του συνόλου A_0 των δραστηριοτήτων του μοντέλου M_0 , στο σύνολο A_1 των δραστηριοτήτων του M_1 . [22][23]

¹⁵Με A^* συμβολίζεται το σύνολο όλων των πεπερασμένων ακολουθιών από στοιχεία του A .

Το επόμενο βήμα είναι ο υπολογισμός των μέτρων ομοιότητας μεταξύ των μοντέλων. Αυτό επιτυγχάνεται συνήθως με συμμετρικές και μη αρνητικές συναρτήσεις απόστασης που αποτυπώνουν το ποσό των διαφορών που παρουσιάζει ένα ζεύγος μοντέλων διεργασιών. [23]

Ένα μέτρο ομοιότητας πρέπει να έχει τις παρακάτω ιδιότητες:

Έστω M_0 και M_1 τα μοντέλα που συγκρίνονται και $(M, dist)$ ένας μετρικός χώρος, όπου M το σύνολο όλων των μοντέλων με ένα μέτρο απόστασης $dist$.

Για να είναι το μέτρο $dist$ μέτρο ομοιότητας πρέπει [22]

- $dist(M_0, M_1) \geq 0$, $\forall M_0, M_1 \in M$, δηλαδή να είναι μη αρνητικό
- $dist(M_0, M_1) = dist(M_1, M_0)$, $\forall M_0, M_1 \in M$, δηλαδή να είναι συμμετρικό
- $dist(M_0, M_1) = 0$, αν και μόνο αν τα μοντέλα M_0, M_1 είναι ίδια ή έχουν το ίδιο σύνολο Ιχνών
- $dist(M_0, M_2) \leq dist(M_0, M_1) + dist(M_1, M_2)$, δηλαδή ικανοποιούν την τριγωνική ανισότητα
- Το μέτρο λαμβάνει υπόψη τόσο τις ομοιότητες όσο και τις διαφορές¹⁶
- Το μέτρο λαμβάνει υπόψιν το μέτρο ομοιότητας μεταξύ δραστηριοτήτων¹⁷
- Το μέτρο ορίζεται για αυθαίρετα μοντέλα διεργασιών (δηλαδή χωρίς την επιβολή περιορισμών όπως το να περιέχουν τα προς σύγκριση μοντέλα βρόγχους ή όχι)
- Το μέτρο μπορεί να υπολογιστεί αποτελεσματικά

Λαμβάνοντας υπόψιν τις ιδιότητες αυτές έχουν αναπτυχθεί διαφορετικές μέθοδοι για τον υπολογισμό μέτρων ομοιότητας. Οι μέθοδοι αυτές μπορούν να ομαδοποιηθούν σε 7 κατηγορίες: «απλοποίηση αλλαγών», «συγχώνευση διαδικασιών», «διευκόλυνση της επαναχρησιμοποίησης», «διαχείριση αποθετηρίων Μοντέλων Διεργασιών», «αυτοματοποίηση της εκτέλεσης διεργασιών», «διασφάλιση της συμμόρφωσης με τα κανονιστικά μοντέλα», και «ανακάλυψη υπηρεσιών» [22], ενώ επικεντρώνονται κυρίως είτε στο γράφημα του μοντέλου διεργασίας είτε τη συμπεριφορά που αποτυπώνεται στο μοντέλο [23].

2.3 Η Ανάλυση των Βασικών Αιτίων

Εφόσον έχει εξορυχθεί ένα Μοντέλο Διεργασιών και αφού έχουν βρεθεί οι περιπτώσεις που στην πραγματικότητα αποκλίνουν από το Μοντέλο είναι σημαντικό να εντοπιστούν οι αι-

¹⁶Η ιδιότητα αυτή περιγράφει ότι για να έχουμε μία καλή εικόνα της ομοιότητας δύο μοντέλων δεν αρκεί να βρούμε πόσες δραστηριότητες είναι κοινές, αλλά και το σύνολο των δραστηριοτήτων των μοντέλων (π.χ. η ύπαρξη 10 κοινών δραστηριοτήτων σε ένα σύνολο 15 δραστηριοτήτων είναι συνεπάγεται μεγαλύτερη ομοιότητα από την ύπαρξη 10 κοινών δραστηριοτήτων σε ένα σύνολο 100 δραστηριοτήτων)

¹⁷Το μέτρο ομοιότητας δραστηριοτήτων $corr : A_0 \times A_1 \rightarrow [0, 1]$ ορίζεται για $a_0 \in A_0$, $a_1 \in A_1$, όπου A_0, A_1 τα σύνολα δραστηριοτήτων των M_0, M_1 αντίστοιχα, με τρόπο ώστε:

- $cor(a_0, a_1) = 1$, αν $map(a_0) = a_1$, $map : M_0 \rightarrow M_1$ και $a_0 = a_1$, δηλαδή οι δραστηριότητες a_0 και a_1 παρουσιάζονται με τον ίδιο ακριβώς τρόπο στα δύο μοντέλα
- $cor(a_0, a_1) \leq 1$, σε κάθε άλλη περίπτωση

τίες που σχετίζονται με τις αποκλίνουσες συμπεριφορές, δηλαδή τις συμπεριφορές που αποκλίνουν από την συνηθισμένη συμπεριφορά του Μοντέλου, αλλά και οι αιτίες που σχετίζονται με τις περιπτώσεις όπου η απόδοση της Διεργασίας είναι χαμηλή, δηλαδή δεν επιτυγχάνεται ο στόχος που έχει τεθεί. Παραδείγματος χάριν όταν μελετάμε Διεργασίες που αφορούν την έγκριση δανείων μία περίπτωση κακής απόδοσης μπορεί να είναι ο μεγάλος χρόνος αναμονής μέχρι την έγκριση ή απόρριψη της αίτησης και αντίστοιχη αιτία μπορεί να είναι ο μεγάλος όγκος πελατών το διάστημα εκείνο. Σκοπός της Ανάλυσης των Βασικών Αιτιών είναι ο εντοπισμός, δηλαδή να εντοπισθούν οι αιτίες εκείνες που προκαλούν οποιαδήποτε μη επιθυμητή συμπεριφορά σε κάποιες Υποθέσεις, προκειμένου να βελτιωθεί η συνολική απόδοση της Διεργασίας. Κάτι τέτοιο είναι ιδιαίτερα απαιτητικό, καθώς δύναται να υπάρχουν παράγοντες οι οποίοι επηρεάζουν έμμεσα την απόδοση μιας Διεργασίας και έτσι είναι δύσκολο να εντοπιστούν [24], ενώ επίσης η ύπαρξη συσχέτισης μεταξύ δύο γεγονότων (που είναι εύκολο να ανιχνευθεί με διάφορες στατιστικές μεθόδους) δεν υποδηλώνει κατ' ανάγκη αιτιότητα. Έχουν αναπτυχθεί διάφορες Μέθοδοι για την Ανάλυση των Βασικών Αιτιών, οι οποίες βασίζονται στις τεχνικές ταξινόμησης, την εξόρυξη κανόνων, την ανακάλυψη υποομάδων κ.α., ωστόσο στις περισσότερες περιπτώσεις δεν λαμβάνεται υπόψιν η διάκριση μεταξύ Συσχέτισης και Αιτιότητας. [25] Επίσης ενώ έχουν αναπτυχθεί πολλές τεχνικές Ανάλυσης Αιτιών που χρησιμοποιούν παραδοσιακές Μεθόδους Στατιστικής, δεν έχουν αναπτυχθεί σε μεγάλο βαθμό μέθοδοι ανάλυσης αιτιών με τεχνικές εξόρυξης δεδομένων.[26] Παρακάτω αναφέρονται Μέθοδοι Ανάλυσης Βασικών Αιτιών με τεχνικές εξόρυξης δεδομένων, οι οποίες λαμβάνουν υπόψιν την διάκριση μεταξύ συσχέτισης και αιτιότητας.

2.3.1 Η Μέθοδος του Hompes

Οι Hompes et all [27] προτείνουν μία Μέθοδο η οποία δέχεται ένα Αρχείο Καταγραφής Γεγονότων και παράγει ένα γράφημα αιτιωδών παραγόντων (Causal Factors) που εξηγούν την απόδοση της Διεργασίας. Η μέθοδος αυτή εντοπίζει σχέσεις αιτιότητας μεταξύ μιας σειράς χαρακτηριστικών επιχειρηματικών διαδικασιών και δεικτών απόδοσης διαδικασιών, όπως η διάρκεια των Υποθέσεων και χρόνος αναμονής δραστηριότητας, χρησιμοποιώντας τον Στατιστικό Έλεγχο του Granger [28], που χρησιμοποιείται για την εύρεση αιτιωδών σχέσεων μεταξύ χρονοσειρών. Η βασική ιδέα της Μεθόδου των Hompes et all είναι ότι οι τιμές που προσδιορίζουν την απόδοση μιας διεργασίας θεωρούνται χρονοσειρές. Ένας παράγοντας λέγεται ότι είναι αιτιώδης για έναν άλλο όταν οι παρελθούσες τιμές αυτού παρέχουν πληροφορίες που μπορούν να βοηθήσουν στην πρόβλεψη του άλλου παράγοντα, πέρα από τις πληροφορίες που περιέχονται στις παρελθούσες τιμές του τελευταίου παράγοντα από μόνες τους. Ο εντοπισμός των αιτιωδών παραγόντων γίνεται σε 3 βήματα. Βήμα 1^ο: Η βάση γεγονότων αναλύεται συστηματικά σε ένα κατευθυνόμενο ακυκλικό γράφημα. Στο γράφημα αυτό κάθε κόμβος αναπαριστά μία συλλογή γεγονότων τα οποία μοιράζονται ορισμένα χαρακτηριστικά της διαδικασίας και μπορούν να θεωρηθούν ως εν δυνάμει αιτιώδεις παράγοντες (π.χ. τύπο της υπόθεσης, το όνομα των δραστηριοτήτων κλπ.). Επίσης οι ακμές του γραφήματος συνδέουν έναν κόμβο-πηγή με έναν κόμβο - στόχο, όταν ο κόμβος - στόχος είναι το αποτέλεσμα περαιτέρω ανάλυσης του κόμβου πηγής χρησιμοποιώντας οποιοδήποτε χαρακτηριστικό της διαδικασίας, δηλαδή το σύνολο των ιδιοτήτων του πρώτου περιλαμβάνεται στις ιδιότητες

του δεύτερου και το σύνολο γεγονότων του δεύτερου είναι υποσύνολο του συνόλου γεγονότων του πρώτου. Βήμα 2^ο: Το γράφημα ανάλυσης (decomposition graph) μετατρέπεται σε ένα γράφημα συμπερίληψης (inclusion graph), οι ακμές του οποίου αντιπροσωπεύουν τις υποψήφιες αιτιώδεις σχέσεις μεταξύ παραγόντων (κόμβοι). Η απόδοση που σχετίζεται με τον κόμβο - πηγή είναι δυνητικός αιτιώδης παράγοντας για την απόδοση που σχετίζεται με τον κόμβο - στόχο. Αρχικά είναι συνδεδεμένα όλα τα ζεύγη κόμβων και έπειτα αφαιρούνται οι ακμές μεταξύ ζευγών κόμβων που έχουν προγονική σχέση στο γράφημα ανάλυσης (η συσχέτιση γεγονότων που σχετίζονται με τον τύπο υπόθεσης x και το όνομα δραστηριότητας y (κόμβος - απόγονος) με τα γεγονότα που σχετίζονται με τον τύπο υπόθεσης x (κόμβος - πρόγονος) δεν έχει νόημα, ενώ αφαιρούνται επίσης κόμβοι και ακμές που είναι άνευ σημασίας από επιχειρηματικής άποψης. Βήμα 3^ο: για κάθε ζεύγος συνδεδεμένων κόμβων στο γράφημα συμπερίληψης, οι τιμές απόδοσης για τα γεγονότα κάθε κόμβου μετατρέπονται σε χρονοσειρές και η αιτιότητα τους ελέγχεται με τον Έλεγχο Granger. Η μετατροπή των γεγονότων σε χρονοσειρές, δηλαδή σε μια χρονικά ταξινομημένη συλλογή πραγματικών τιμών, γίνεται μέσω της εφαρμογής μία συνάρτησης απόδοσης (Performance function) σε ένα σύνολο γεγονότων.

Ορισμός 2.17 (Συναρτηση Απόδοσης Performance function). Έστω I το σύνολο δεικτών απόδοσης. Συνάρτηση Απόδοσης είναι μία συνάρτηση $\vartheta \in I \rightarrow (E \rightarrow \mathbb{R} \times \mathcal{T})$, δηλαδή για κάθε δείκτη απόδοσης $i \in I$, η $\vartheta(i)$ είναι μία μερική συνάρτηση¹⁸, η οποία απεικονίζει γεγονότα $e \in E$ σε χρονικές πραγματικές τιμές $(r, t) \in \mathbb{R} \times \mathcal{T}$. Η έκφραση $\vartheta_i(e) = (r, t)$ συμβολίζει ότι η απόδοση του γεγονότος e είναι $r \in \mathbb{R}$ και η χρονική σήμανση του γεγονότος e είναι η $t \in \mathcal{T}$. Εάν $e \in E \setminus \text{dom}(\vartheta(i))$ ¹⁹, δηλαδή το γεγονός e δεν έχει τιμή για τον δείκτη απόδοσης i , τότε γράφουμε $\vartheta(i) = \perp$.

Για να γίνει πιο κατανοητός ο ορισμός 2.17 δίνεται ο τύπος της συνάρτησης απόδοσης για τον δείκτη που αφορά την διάρκεια μιας υπόθεσης:

$\vartheta_{\text{διάρκεια υπόθεσης}}(e) = (\max\{\pi_t(e') \mid e' \in E \cap \{\pi_c(e') = \pi_c(e)\}\} - \min\{\pi_t(e') \mid e' \in E \cap \{\pi_c(e') = \pi_c(e)\}\}, \max\{\pi_t(e') \mid e' \in E \cap \{\pi_c(e') = \pi_c(e)\}\})$, όπου $\pi_t(e) = u$ συμβολίζει ότι το γεγονός e έχει την ιδιότητα t και η τιμή της ιδιότητας αυτής είναι u και $\pi_t = \pi_{\text{time}}$, $\pi_c = \pi_{\text{case}}$.

Ο τύπος της $\vartheta_{\text{διάρκεια υπόθεσης}}(e)$ δηλώνει ότι η $\vartheta(\text{διάρκεια υπόθεσης})$ απονέμει στο γεγονός e την τιμή $\max\{\pi_t(e') \mid e' \in E \cap \{\pi_c(e') = \pi_c(e)\}\} - \min\{\pi_t(e') \mid e' \in E \cap \{\pi_c(e') = \pi_c(e)\}\}$, η οποία στην πραγματικότητα είναι η διάρκεια της υπόθεσης, ως διαφορά μεταξύ της μέγιστης χρονικής στιγμής στην οποία εκτελείται κάποιο γεγονός της υπόθεσης c και της ελάχιστης χρονικής στιγμής στην οποία εκτελείται κάποιο γεγονός της υπόθεσης c , και η αντίστοιχη χρονική σήμανση είναι η χρονική στιγμή λήξης της υπόθεσης (δηλαδή η μέγιστη χρονική στιγμή στην οποία εκτελείται κάποιο γεγονός της υπόθεσης c). Με άλλα λόγια $\vartheta_{\text{διάρκεια υπόθεσης}}(e) = (\text{διάρκεια υπόθεσης}, \text{στιγμή ολοκλήρωσης υπόθεσης})$

Με τον τρόπο αυτό μπορούμε έχουμε μία χρονοσειρά $\{\text{διάρκεια}_1, \text{διάρκεια}_2, \text{διάρκεια}_3, \dots\}$ ταξινομημένη με βάση την στιγμή ολοκλήρωσης (δηλαδή στιγμή ολοκλήρωσης₁ < στιγμή ολοκλήρωσης₂ < στιγμή ολοκλήρωσης₃ < ...).

¹⁸Η $\vartheta(i)$ είναι μία μερική συνάρτηση, δηλαδή υπάρχουν στοιχεία του E τα οποία δεν αντιστοιχίζονται μέσω της ϑ σε στοιχεία του $\mathbb{R} \times \mathcal{T}$

¹⁹Η έκφραση $e \in E \setminus \text{dom}(\vartheta(i))$ δηλώνει ότι το e δεν απεικονίζεται μέσω της $\vartheta(i)$ σε κάποιο στοιχείο του $\mathbb{R} \times \mathcal{T}$

Στο σημείο αυτό μπορεί να πραγματοποιηθεί ο έλεγχος αιτιότητας του Granger. Σύμφωνα με αυτόν μία χρονοσειρά S είναι αιτία μίας άλλης χρονοσειράς S' αν οι προηγούμενες τιμές του S βοηθούν στην πρόβλεψη των μελλοντικών τιμών του S' καλύτερα από ό,τι οι παρελθοντικές τιμές του S' μπορούν να προβλέψουν τον εαυτό τους. Για την διεξαγωγή του ελέγχου αρχικά προσαρμόζεται ένα γραμμικό αυτοπαλίνδρομο μοντέλο μίας μεταβλητής για το S' ως εξής:

$$s'_{tu} = \sum_{k=1}^L a'_k \cdot s'_{tu-k} + \epsilon'_t,$$

Στην συνέχεια προσαρμόζεται ένα γραμμικό αυτοπαλίνδρομο μοντέλο 2 μεταβλητών για το S' :

$$s'_{tb} = \sum_{k=1}^L a_k \cdot s'_{tb-k} + \sum_{k=1}^L b_k \cdot s_{t-k} + \epsilon_t,$$

Στους παραπάνω τύπους L είναι η υστέρηση του ελέγχου Granger, $t = L + 1, \dots, |S'|$, a , a' , b διανύσματα παραμέτρων και ϵ' τα σφάλματα, τα οποία μπορούν να υπολογιστούν με εκτιμήτριες ελάχιστων τετραγώνων.

Στην πραγματικότητα στο μοντέλο μίας μεταβλητής s'_{tu} οι τιμές του S' περιγράφονται μόνο από παρελθοντικές του ίδιου του S' , ενώ στο μοντέλο δύο μεταβλητών οι τιμές του S' περιγράφονται τόσο από τι παρελθοντικές τιμές του όσο και από τιμές του s . Τέλος, ο έλεγχος γίνεται με βάση το στατιστικό $G = \frac{(\epsilon' - \epsilon)L}{\epsilon \sqrt{(|S'| - 2L)}}$, το οποίο ακολουθεί την κατανομή F με L και $|S'| - 2L$ βαθμούς ελευθερίας. Μεγάλη τιμή του G υποδηλώνει ότι η πληροφορία του παρελθόντος στο S είναι χρήσιμη για την πρόβλεψη μελλοντικών τιμών στο S' .

Ένα σημαντικό πρόβλημα της Μεθόδου αυτής είναι ότι εξ ορισμού η αιτιότητα κατά Granger δεν λαμβάνει υπόψιν τις λανθάνουσες επιδράσεις και δεν καταγράφει τις στιγμιαίες και μη γραμμικές αιτιώδεις σχέσεις.[27]

2.3.2 Η Μέθοδος των Van der Aalst και Qafari

[29] Οι Van de Aalst και Qafari προτείνουν μία Μέθοδο που βασίζεται στον Εμπλουτισμό του Αρχείου Καταγραφής Γεγονότων με χαρακτηριστικά τα οποία προκύπτουν με υπολογισμούς μεταξύ υπάρχοντων χαρακτηριστικών ή από εξωτερικές πηγές και στην χρήση των αντιφατικών περιπτώσεων (Counterfactual instances). Το πρώτο βήμα της μεθόδου είναι ο εμπλουτισμός του Αρχείου Καταγραφής Γεγονότων.[29] Το Εμπλουτισμένο Αρχείο Καταγραφής Γεγονότων είναι το Αρχείο Καταγραφής Γεγονότων που περιέχει τα Γεγονότα του Αρχικού Αρχείου Καταγραφής, κάθε ένα από τα οποία έχει τα χαρακτηριστικά τα οποία υπήρχαν στο αρχικό Αρχείο Καταγραφής Γεγονότων (π.χ. ένδειξη χρόνου), χαρακτηριστικά που προκύπτουν ως το αποτέλεσμα υπολογισμών μεταξύ των αρχικών χαρακτηριστικών μέσω μιας συνάρτησης παραγωγής (π.χ. χρονική διάρκεια μεταξύ δύο γεγονότων) και ένα σύνολο συσχετιζόμενων χαρακτηριστικών των οποίων η τιμή μπορεί να ληφθεί από εξωτερικές πηγές με βάση κάποια χαρακτηριστικά - κλειδιά (key - attributes) του αρχικού συνόλου χαρακτηριστικών μέσω μίας συσχετιστικής συνάρτησης. [26]. Μία σημαντική παραδοχή είναι ότι ένα αίτιο προηγείται χρονικά του αποτελέσματός του. Συνεπώς δεν μπορούμε να αναζητούμε την αιτία ενός γεγονότος σε γεγονότα που έπονται αυτού. Για τον λόγο αυτό χρησιμοποιούμε την έννοια της Κατάστασης (situation), δηλαδή ένα κατάλληλο πρόθεμα (prefix) του Ίχνους και την έννοια του Συνόλου Καταστάσεων ενός Αρχείου Καταγραφής Γεγονότων, δηλαδή το σύνολο των καταστάσεων που παράγονται από όλα τα Ίχνη του Αρχείου, παραδείγματος χάριν

ένα υποσύνολο του συνόλου καταστάσεων μπορεί να είναι οι καταστάσεις που τελειώνουν σε γεγονός με κάποιο συγκεκριμένο όνομα δραστηριότητας και μία κατάσταση του υποσυνόλου αυτού είναι το μέρος ενός Ίχνους που περιλαμβάνει τα γεγονότα μέχρι και το γεγονός με το συγκεκριμένο όνομα δραστηριότητας. Σε ένα Αρχείο Καταγραφής Γεγονότων μπορεί να υπάρχουν χαρακτηριστικά τόσο στα Γεγονότα όσο και στα Ίχνη. Στο σημείο αυτό θα εισάγουμε επίσης την έννοια του Γνώρισματος Κατάστασης (Situation Feature). Ένα Γνώρισμα Κατάστασης $γκ_{\text{χαρακτηριστικό}}$ αναφέρεται σε ένα χαρακτηριστικό (γεγονότος ή Ίχνους) και σε ένα όνομα δραστηριότητας ή σε ένα Ίχνος, π.χ. το $γκ_{\text{Διάρκεια}\Delta 1}$ είναι το γνώρισμα κατάστασης για την διάρκεια της δραστηριότητας $\Delta 1$ και $γκ_{\text{εμπλεκόμενοι}}$ είναι το γνώρισμα κατάστασης για το χαρακτηριστικό ίχνους “εμπλεκόμενοι”.

Παράδειγμα 2.4. Έστω το Αρχείο Καταγραφής Γεγονότων του Πίνακα 2.5 εμπλουτισμένο με το χαρακτηριστικό Ίχνους “Παλαιότητα Πελάτη”:

Αναγνωριστικό Γεγονότος	Όνομα Πελάτη	Δραστηριότητα	Χρονική Σήμανση	Υπάλληλος	Παλαιότητα Πελάτη
Γ1	A	Υποβολή της Αίτησης	21-04-2023:11.34	Υπάλληλος 1	Κάτω από 1 έτος
Γ2	B	Υποβολή της Αίτησης	21-04-2023:12.34	Υπάλληλος 1	2 με 10 έτη
Γ3	A	Αποδοχή της Αίτησης	21-04-2023:12.40	Υπάλληλος 2	Κάτω από 1 έτος
Γ4	B	Αποδοχή της Αίτησης	21-04-2023:13.34	Υπάλληλος 3	2 με 10 έτη
Γ5	Γ	Υποβολή της Αίτησης	21-04-2023:14.34	Υπάλληλος 3	2 με 10 έτη
Γ6	B	Αναζήτηση του Αρχείου	21-04-2023:14.55	Υπάλληλος 2	2 με 10 έτη
Γ7	Δ	Υποβολή της Αίτησης	22-04-2023:11.34	Υπάλληλος 1	πάνω από 10 έτη
Γ8	Γ	Απόρριψη της Αίτησης	22-04-2023:13.00	Υπάλληλος 2	2 με 10 έτη
Γ9	Γ	Ενημέρωση του Αιτούντος	22-04-2023:13.14	Υπάλληλος 3	2 με 10 έτη
Γ10	A	Αναζήτηση του Αρχείου	22-04-2023:14.04	Υπάλληλος 1	Κάτω από 1 έτος
Γ11	Δ	Αποδοχή της Αίτησης	30-04-2023:12.05	Υπάλληλος 1	πάνω από 10 έτη
Γ12	A	Ενημέρωση του Αιτούντος	30-04-2023:12.05	Υπάλληλος 2	Κάτω από 1 έτος
Γ13	Δ	Αναζήτηση του Αρχείου	30-04-2023:12.09	Υπάλληλος 3	πάνω από 10 έτη
Γ14	B	Ενημέρωση του Αιτούντος	30-04-2023:13.14	Υπάλληλος 1	2 με 10 έτη
Γ15	Δ	Ενημέρωση του Αιτούντος	30-04-2023:14.30	Υπάλληλος 1	πάνω από 10 έτη

Πίνακας 2.5: Αρχείο Καταγραφής Γεγονότων της Υπηρεσίας

Το υποσύνολο καταστάσεων που λήγουν σε δραστηριότητα με όνομα ‘Αποδοχή της Αίτησης’ περιέχει τα γεγονότα με Δραστηριότητες ‘Υποβολή Αίτησης’ και ‘Αποδοχή Αίτησης’, ενώ μία κατάσταση του υποσυνόλου αυτού που εξάγεται από το Ίχνος της Υπόθεσης που αφορά τον Πελάτη A είναι τα Γεγονότα με Αναγνωριστικό Γεγονότος Γ1 και Γ3 (έστω κατάσταση κ_1). Επίσης η τιμή της κ_1 το Γνώρισμα Κατάστασης $γκ_{\text{ΥπΑιΥπαβ}}$ (που δηλώνει τον Υπάλληλο στον οποίο Υποβλήθηκε η Αίτηση) είναι ‘Υπάλληλος 1’, ενώ η τιμή του Γνωρισματος Κατάστασης $γκ_{\text{παλαιότητα}}$ (που αναφέρεται στην παλαιότητα του πελάτη) είναι ‘Κάτω από 1 έτος’.

Το Γνώρισμα Κατάστασης - Στόχος (Target Situation Feature) ορίζει το πρόβλημα το οποίο μας απασχολεί και ένα σύνολο περιγραφικών γνωρισμάτων κατάστασης αποτελούν τις πιθα-

νές αιτίες του. Το γνώρισμα κατάστασης - στόχος και τα περιγραφικά γνωρίσματα κατάστασης αποτελούν ένα σχέδιο εξαγωγής γνωρισμάτων κατάστασης (situation feature extraction plan) το οποίο συμβολίζεται με SF. Στιγμιότυπο (instance) i είναι ένα σημείο δεδομένων το οποίο απεικονίζει τις τιμές των γνωρισμάτων κατάστασης κάποιας κατάστασης του SF. Στο παραπάνω παράδειγμα μπορούμε να πάρουμε ως Γνώρισμα Κατάστασης - Στόχο την διάρκεια Αναζήτησης Αρχείου, δηλαδή το $\gamma_{\text{διάρκειαΑναζήτηση}}$ και ως περιγραφικά γνωρίσματα κατάστασης να χρησιμοποιηθούν τα $\gamma_{\text{υπάλληλοςΥποβολή}}$, $\gamma_{\text{τελάτηςΠαλαιότητα}}$, $\gamma_{\text{διάρκειαΥποβολή}}$ και τότε $SF_{\text{διάρκειαΑναζήτηση}} = \gamma_{\text{υπάλληλοςΥποβολή}}$, $\gamma_{\text{τελάτηςΠαλαιότητα}}$, $\gamma_{\text{διάρκειαΥποβολή}}$, $\gamma_{\text{διάρκειαΑναζήτηση}}$ και ένα στιγμιότυπο που παράγεται από την κ_1 μέσω του $SF_{\text{διάρκειαΑναζήτηση}}$ είναι το ($\gamma_{\text{υπάλληλοςΥποβολή}}$, Υπάλληλος 1), ($\gamma_{\text{τελάτηςΠαλαιότητα}}$, κάτω από 1 έτος), ($\gamma_{\text{διάρκειαΥποβολή}}$, 66), ($\gamma_{\text{διάρκειαΑναζήτηση}}$, 11401)²⁰. Τέλος το σύνολο των τροποποιησιμων γνωρισμάτων κατάστασης ονομάζονται Αξιοποιήσιμα Γνωρίσματα Κατάστασης (actionable situation features και συμβολίζεται ASF).

Ορισμός 2.18 (Σύνολο Αντιφατικών Αιτιολογήσεων (Set of Counterfactual Explanation)). *Έστω i ένα στιγμιότυπο που απεικονίζει μία κατάσταση για κάποιο Σχέδιο Εξαγωγής Γνωρισμάτων Κατάστασης SF. Έστω ότι η τιμή του γνωρίσματος κατάστασης - στόχου του i είναι μη επιθυμητή με βάση κάποιο κατόφθι t που έχει οριστεί. Ένα σύνολο Αιτιολογήσεων για το i είναι ένα σύνολο διακριτών περιπτώσεων που βρίσκονται κοντά στο i , διαφέρουν από το i κατά ένα υποσύνολο του συνόλου των τροποποιήσιμων γνωρισμάτων κατάστασης και έχουν ένα επιθυμητό αποτέλεσμα για το γνώρισμα κατάστασης-στόχος.*

Οι Αντιφατικές Αιτιολογήσεις παράγονται μέσα από μία διαδικασία κατά την οποία αρχικά παράγονται υποψήφιος αιτιολογήσεις αποδίδοντας τιμές στα γνωρίσματα κατάστασης του ASF Έπειτα υπολογίζεται η επίδραση της αντικατάστασης των τιμών των γνωρισμάτων κατάστασης στο συγκεκριμένο στιγμιότυπο χρησιμοποιώντας το Μοντέλο διαρθρωτικών εξισώσεων (Structural Equation Model (SEM)), δηλαδή ένα σύνολο εξισώσεων που καθορίζουν πώς μπορούν να δημιουργηθούν οι κατανομές παρατήρησης και παρέμβασης. Με πιο απλά λόγια, μέσα από ένα σύνολο εξισώσεων και με την βοήθεια μοντέλων προβλέπουμε την τιμή του γνωρίσματος κατάστασης - στόχου για κάθε μία από τις υποψήφιος Αιτιολογήσεις καταλήγοντας έτσι σε ένα σύνολο στιγμιότυπων που ονομάζονται Αντιφατικά Στιγμιότυπα (Counterfactual Instances). Επίσης με την χρήση τεχνικών βελτιστοποίησης, τα στιγμιότυπα αυτά μπορούν να έρθουν όσο το δυνατό πιο κοντά στο υπό εξέταση στιγμιότυπο. Τέλος τα αντιφατικά στιγμιότυπα που έχουν προκύψει ως αποτέλεσμα της παραπάνω διαδικασίας ταξινομούνται με βάση την απόσταση τους από το υπό εξέταση στιγμιότυπο και μετατρέπονται σε ένα σύνολο εξηγήσεων. [29]

2.3.3 Ο Αλγόριθμος ΑΙΤΙΑ - PM

Οι van Houdt et al [25] προτείνουν τον Αλγόριθμο ΑΙΤΙΑ-PM, ο οποίος χρησιμοποιεί την πιθανολογική χρονική λογική (probabilistic temporal logic) για την δόμηση υποθέσεων σχετικά με τις σχέσεις αιτιότητας, ενώ βασίζεται στον Αλγόριθμο του Kleinberg επεκτείνοντάς

²⁰ Η διάρκεια υπολογίζεται σε λεπτά της ώρας μέσω του χαρακτηριστικού 'Χρονική Σήμανση' και με την παραδοχή ότι για κάθε πελάτη μία δραστηριότητα λήγει την στιγμή που συμβαίνει η επόμενη.

τον στην εξόρυξη διεργασιών. Για να οριστεί επαρκώς η έννοια της αιτιότητας γίνονται οι εξής παραδοχές: (1) το αίτιο πρέπει να προηγείται χρονικά του αποτελέσματος, (2) η ύπαρξη ενός αιτίου πρέπει να αυξάνει την πιθανότητα του αποτελέσματος και (3) κάθε Υπόθεση ορίζεται από τα Γεγονότα που δύνανται να είναι αιτίες των αποτελεσμάτων μέσα σε αυτή την Υπόθεση. Τα γεγονότα που πληρούν τις προϋποθέσεις (1) ως (3) ονομάζονται εκ πρώτης όψεως αιτίες (prima facie causes). Η Πιθανολογική Χρονική Λογική επιτρέπει τη συλλογιστική σχετικά με την πιθανότητα να συμβεί ένα γεγονός εντός ενός συγκεκριμένου χρονικού διαστήματος. Ο Αλγόριθμος ΑΙΤΙΑ - ΡΜ χρησιμοποιεί την Πιθανολογική Χρονική Λογική ως την γλώσσα στην οποία ορίζεται η υπόθεση που πρόκειται να ελεγχθεί. Κάθε υπόθεση περιλαμβάνει έναν λογικό τύπο ο οποίος περιγράφει τα χρονικά διάστημα και την πιθανότητα που έχει ένα εν δυνάμει αίτιο να προκαλέσει ένα αποτέλεσμα. Μία υπόθεση Πιθανολογικής Χρονικής Λογικής διατυπώνεται ως τύπος με την εξής μορφή: $c \rightsquigarrow_{\geq p}^{\geq r, \leq s} e$, το οποίο εξηγείται ως εξής: η πιθανότητα η αιτία c να προκαλεί το αποτέλεσμα e στο χρονικό πλαίσιο μετράζω r και s είναι τουλάχιστον p . Το πρώτο βήμα του Αλγόριθμου είναι η προετοιμασία των δεδομένων, δηλαδή να οριστεί ποια δεδομένα θα τεθούν υπό επεξεργασία και με ποια δεδομένα πρέπει να εμπλουτιστεί το Αρχείο Κατγράφης, ανάλογα με τις υποθέσεις που πρόκειται να ελεγχθούν. Έπειτα ακολουθεί ο καθορισμός των Υποθέσεων. Στο σημείο αυτό διατυπώνονται όλες οι πιθανές υποθέσεις σχετικά με το αν ένα γεγονός c προκαλεί κάποιο αποτέλεσμα e μέσα στα χρονικά πλαίσια $[r, s]$. Εφόσον έχουν οριστεί όλες οι πιθανές σχέσεις που μπορούν να αποτελούν σχέση αίτιου - αποτελέσματος πρέπει να αλεγχθεί αν πληρούνται οι προϋποθέσεις για τις εκ πρώτης όψεως αιτίες. Αυτό ελέγχεται με την εξής σχέση $P(e|c) > P(e)$ (δηλαδή η πιθανότητα να συμβεί το e εφόσον έχει συμβεί το c είναι μεγαλύτερη από την πιθανότητα να συμβεί το e). Σημειώνεται ότι $P(e) = \frac{\#e}{\text{Αριθμός Γεγονότων}}$ και $P(e|c) = \frac{\#e \wedge c}{\#e}$ (με $e \wedge c$ συμβολίζεται η πραγματοποίηση των γεγονότων e και c), όπου στον υπολογισμό του $\#e \wedge c$ προσμετρώνται μόνο οι περιπτώσεις στις οποίες τα e και c ανήκουν στην ίδια Υπόθεση και το c συμβαίνει πριν το e . Στο σημείο αυτό πρέπει να βρεθούν οι πραγματικές αιτίες και για τον σκοπό αυτό εισάγεται το μέτρο αιτιότητας ϵ_{avg} του Kleinberg, το οποίο περιγράφει την μέση μεταβολή της πιθανότητας του αποτελέσματος e δεδομένης της παρουσίας της αιτίας c κρατώντας κάποιον παράγοντα x σταθερό, ο οποίος είναι και αυτός εκ πρώτης όψεως αιτία. Συγκεκριμένα $\epsilon_{avg}(c, e) = \frac{\sum_{x \in X|c} \epsilon_x(c, e)}{|X|c|}$, όπου $\epsilon_x(c, e) = P(e|c \wedge x) - P(e|\neg c \wedge x)$ και X το σύνολο των εκ πρώτης όψεως αιτιών του e , ενώ επίσης $P(e|c \wedge x) = \frac{\#(e \wedge c \wedge x)}{\#(c \wedge x)}$ και $P(e|\neg c \wedge x) = \frac{\#(e \wedge \neg c \wedge x)}{\#(\neg c \wedge x)}$, όπου για τον υπολογισμό των $\#(e \wedge c \wedge x)$ και $\#(e \wedge \neg c \wedge x)$ προσμετρώνται μόνο οι περιπτώσεις στις οποίες τα e , c (ή αντίστοιχα $\neg c$) και x ανήκουν στην ίδια Υπόθεση και το e συμβαίνει μετά το $c \wedge x$ (ή $\neg c \wedge x$ αντίστοιχα). Ο στατιστικός έλεγχος γίνεται με την χρήση των ποσοστών ψευδούς ανακάλυψης (f false discovery rates (FDR)) ως εξής: αρχικά υπολογίζονται οι τιμές z του ελέγχου από τον τύπο $z = \frac{\epsilon_{avg} - \mu}{\sigma}$, όπου μ και σ ο μέσος και η τυπική απόκλιση της κατανομής των ϵ_{avg} , έπειτα προσαρμόζεται ένα μοντέλο στις παρατηρούμενες τιμές z και τέλος καθορίζεται το FDR του z . Οι σχέσεις για τις οποίες το FDR είναι κάτω από μία τιμή η οποία έχει οριστεί ως κατώφλι θεωρούνται σημαντικές αιτίες. [25]

Κεφάλαιο **3**

Εφαρμογή των Τεχνικών Εξόρυξης Διεργασιών σε Δεδομένα από τον Τραπεζικό Τομέα

Στο Κεφάλαιο αυτό παρουσιάζεται το πρόβλημα με το οποίο θα ασχοληθούμε. Με αφετηρία ένα Αρχείο Καταγραφής Γεγονότων που αφορά Δεδομένα για την διαδικασία έγκρισης δανείου σε μία τράπεζα¹ εφαρμόζονται ορισμένες μέθοδοι εξόρυξης διεργασιών και αναλύονται στατιστικά που αφορούν την διεργασία. Στο πρώτο μέρος της επεξεργασίας των δεδομένων με την βοήθεια του εργαλείου ProM εφαρμόζονται μέθοδοι για την ανακάλυψη του Μοντέλου και τον Έλεγχο Προσαρμογής, ενώ στο δεύτερο μέρος με την βοήθεια του εργαλείου QPR Process Analyzer αναλύεται περαιτέρω η Διεργασία.

3.1 Τα Δεδομένα του Προβλήματος

Το Αρχείο Καταγραφής Γεγονότων το οποίο πρόκειται να αναλυθεί περιέχει 89.924 Γεγονοτα τα οποία αφορούν 14.032 Υποθέσεις. Για κάθε Γεγονός καταγράφονται τα εξής:

- Αναγνωριστικό Υπόθεσης - Case ID.
- Τύπος Γεγονότος - Event Type
- Χρονική Σήμανση Έναρξης Γεγονότος - Start Time
- Μέθοδος Εισαγωγής Γεγονότος - Automation

Επιπλέον υπάρχουν διαθέσιμα και ορισμένα χαρακτηριστικά που αφορούν τις Υποθέσεις. Για την Εξόρυξη Διεργασιών αρκεί το Αρχείο Καταγραφής Γεγονότων, ωστόσο εφόσον υπάρχουν διαθέσιμα και περαιτέρω δεδομένα είναι χρήσιμο να αξιοποιηθούν. Τα δεδομένα αυτά βρίσκονται συγκεντρωμένα σε ένα Αρχείο Καταγραφής Υποθέσεων, όπου κάθε γραμμή αναφέρεται σε μία Υπόθεση και καταγράφονται τα εξής:

- Αναγνωριστικό Υπόθεσης - Case ID.
- Επιχειρηματική Μονάδα - Business Unit
- Κόστος- Cost

¹Το αρχείο βρέθηκε ως demo στο περιβάλλον του εργαλείου QPR Process Analyzer

- Αιτία Απόρριψης Δανείου - Loan Rejection Reason
- Κατάσταση Δανείου - Loan Status
- Τοποθεσία - Location
- Μήνας - Month
- Προϊόν - Product
- Συμφωνία επιπέδου υπηρεσιών - SLA
- Χρήστης - User

Στις Εικόνες 3.1 και 3.2 φαίνονται οι πρώτες γραμμές του Αρχείου Καταγραφής Γεγονότων και του Αρχείου Καταγραφής Υποθέσεων.

Case ID	Event Type	Start Time	Automation
16613129	New Loan Application	07/05/2019 20:48:31	Automatic
16613129	Process Loan Application	07/08/2019 00:29:04	Manual
16613129	B2C First Approval	07/10/2019 21:03:06	Manual
16613129	Loan Rejected	07/11/2019 22:49:56	Manual
16613139	New Loan Application	07/05/2019 21:33:27	Manual
16613139	Process Loan Application	07/05/2019 22:30:06	Manual
16613139	B2C First Approval	07/05/2019 23:10:29	Manual
16613139	Loan Approved	07/06/2019 00:12:11	Manual
16613139	Manual Check for Valid Documentation	07/07/2019 23:35:06	Manual
16613139	Process Loan to Customer Account	07/08/2019 00:38:31	Automatic
16613149	New Loan Application	07/05/2019 22:02:05	Manual
16613149	Process Loan Application	07/05/2019 22:36:16	Manual
16613149	Application Changed	07/05/2019 22:56:14	Manual
16613149	Process Loan Application	07/05/2019 23:37:35	Manual
16613149	B2C First Approval	07/05/2019 23:41:45	Manual
16613149	Loan Approved	07/06/2019 00:00:38	Automatic

Εικόνα 3.1: Αρχείο Καταγραφής Γεγονότων

A	B	C	D	E	F	G	H	I	J
Case ID	Business Unit	Cost	Loan rejection reason	Loan Status	Location	Month	Product	SLA	User
2	16612500 Merchant	3312 N/A		Loan approved and processed	Hong Kong	2019/10	Boat/ship loan (15%)	1.In Target	Alexander Serrano
3	16612501 Bank branch	11200 Not enough collateral		Loan rejected	Tyler	2019/06	Consumer loan (ZZ1)	2.Out of Target	Antwan Boyle
4	16612503 Merchant	544 N/A		Loan approved and processed	Hong Kong	2019/06	Consumer loan (XX1)	2.Out of Target	Angeline Yu
5	16612508 Merchant	2704 N/A		Loan application in progress	Hong Kong	2019/06	Boat/ship loan (15%)	1.In Target	Ronan Benson
6	16612510 Bank branch	8000 Poor credit score		Loan rejected	Frisco	2019/06	Consumer loan (ZZ1)	1.In Target	Matthias Freeman
7	16612511 Bank branch	11200 N/A		Loan approved and processed	Sevilla	2019/06	Consumer loan (ZZ1)	2.Out of Target	Kierra Fields
8	16612512 Bank branch	5120 N/A		Loan approved and processed	Toledo	2019/06	Consumer loan (ZZ1)	2.Out of Target	Kamora Shields
9	16612516 Merchant	2877 Poor credit score		Loan rejected	Hong Kong	2019/06	Consumer loan (XX1)	2.Out of Target	Sarah Valdez
10	16612517 Merchant	640 N/A		Loan approved and processed	Colorado Springs	2019/06	Immediate loan (XX%)	2.Out of Target	Sarah Valdez
11	16612520 Bank branch	2400 Other/unclassified		Loan rejected	Kishinev	2019/06	Consumer loan (ZZ1)	2.Out of Target	Bruce Ali
12	16612521 Bank branch	9600 Poor credit score		Loan rejected	Orange	2019/06	Consumer loan (XX1)	2.Out of Target	Justine Cantu
13	16612525 Bank branch	8000 Too risky industry		Loan rejected	Denton	2019/06	Consumer loan (XX1)	1.In Target	Demarcus Escobar
14	16612529 Merchant	8000 N/A		Loan approved and processed	Glasgow	2019/06	Car loan used (12%)	2.Out of Target	Liliana Sawyer
15	16612533 Bank branch	4320 N/A		Loan approved and processed	Essen	2019/06	Consumer loan (XX1)	2.Out of Target	Willow Barry
16	16612535 Bank branch	12800 Not enough cashflow		Loan rejected	Rotterdam	2019/06	Consumer loan (ZZ1)	2.Out of Target	Aidan Suarez
17	16612541 Bank branch	7200 N/A		Loan approved and processed	Sioux Falls	2019/06	Consumer loan (ZZ1)	2.Out of Target	Mason Hamilton
18	16612544 Bank branch	16000 Poor credit score		Loan rejected	Huntsville	2019/06	Consumer loan (ZZ1)	2.Out of Target	Noe Moody
19	16612546 Bank branch	6400 N/A		Loan approved and processed	Poznan	2019/06	Consumer loan (ZZ1)	2.Out of Target	Tyson Hicks
20	16612549 Merchant	5600 N/A		Loan approved and processed	Reno	2019/06	Car loan used (12%)	2.Out of Target	Serenity Graves
21	16612553 Bank branch	8000 Poor credit score		Loan rejected	Helsinki	2019/06	Consumer loan (ZZ1)	2.Out of Target	Vanessa Townsend
22	16612554 Bank branch	6400 N/A		Loan approved and processed	Marseille	2019/06	Consumer loan (XX1)	2.Out of Target	Eliza Graves
23	16612556 Merchant	4000 N/A		Loan approved and processed	Bremen	2019/06	Car loan used (12%)	2.Out of Target	Danica Mcpherson

Εικόνα 3.2: Αρχείο Καταγραφής Υποθέσεων

Πέρα από τα δύο Αρχεία Καταγραφής δεν είναι διαθέσιμη καμία άλλη πληροφορία για την διεργασία. Στο εξής θα εξαγάγουμε πληροφορίες για τα βήματα της διεργασίας, τους αντίστοιχους χρόνους, για στατιστικά που αφορούν την διεκπεραίωση της διεργασίας και για τυχόν δυσλειτουργίες αυτής.

3.2 Εργαλείο ProM

Το ProM (Process Mining framework) [30] είναι ένα πλαίσιο ανοικτού κώδικα για αλγορίθμους εξόρυξης διεργασιών που αναπτύχθηκε στο Eindhoven University of Technology και παρέχει στους χρήστες και τους προγραμματιστές των αλγορίθμων εξόρυξης διεργασιών μια πλατφόρμα που είναι εύκολη στη χρήση την επέκτασή της. Το ProM ενσωματώνει τη λειτουργικότητα πολλών υφιστάμενων εργαλείων εξόρυξης διεργασιών και παρέχει πολλά πρόσθετα εργαλεία εξόρυξης διεργασιών.[31] Επικεντρώνεται σε τρεις βασικούς άξονες, τα αντικείμενα δεδομένων (Data Objects), τα πρόσθετα (plugins) και τους οπτικοποιητές (visualizers), τα οποία οργανώνονται σε πακέτα τα οποία μπορεί να εγκαταστήσει ο χρήστης. [32] Τα Αρχεία Καταγραφής εισάγονται στο περιβάλλον του ProM με την μορφή αρχείων XML και στην συνέχεια η επεξεργασία τους μπορεί να γίνει μέσω των πρόσθετων εργαλείων που προσφέρονται στον χρήστη. Συγκεκριμένα τα πρόσθετα εργαλεία χωρίζονται σε 5 κατηγορίες: τα εργαλεία εξόρυξης, τα οποία υλοποιούν αλγόριθμους ανακάλυψης μοντέλων, τα εργαλεία εξαγωγής, μέσω των οποίων μπορούν να εξαχθούν και να αποθηκευτούν αρχεία που παράγονται στο περιβάλλον του ProM(π.χ. γραφήματα), τα εργαλεία εισαγωγής, μέσω των οποίων μπορούν να φορτωθούν στο σύστημα αντικείμενα, τα εργαλεία ανάλυσης, τα οποία υλοποιούν αναλύσεις για τα αποτελέσματα εξορύξεων (π.χ. έλεγχος προσαρμογής, επεξεργασία δικτύων Petri κ.α.) και τέλος τα εργαλεία μετατροπής, τα οποία υλοποιούν μετατροπές αρχείων από μία μορφή σε άλλη. Κάθε μία από τις 5 κατηγορίες πρόσθετων περιλαμβάνει έναν αριθμό εργαλείων, ενώ υπάρχει επίσης η δυνατότητα κάποιος χρήστης να προσθέσει ένα νέο εργαλείο, όπως για παράδειγμα έναν νέο αλγόριθμο ανακάλυψης μοντέλου. Σκοπός είναι να δίνεται η δυνατότητα σύγκρισης διαφορετικών μεθόδων και αλγορίθμων σε ένα διαδραστικό και φιλικό προς τον χρήστη περιβάλλον. [31]

3.3 Εργαλείο QPR Process Analyzer

Το QPR Process Analyzer είναι ένα λογισμικό το οποίο έχοντας ως αφετηρία ένα Αρχείο Καταγραφής Γεγονότων παρέχει ένα ακριβές Διάγραμμα Ροής της Διεργασίας ως έχει και ταυτόχρονα παρέχει και άλλες δυνατότητες για απεικόνιση των δεδομένων από διαφορετικές οπτικές γωνίες, ενώ η λεπτομέρεια και η έκταση της ανάλυσης εξαρτάται κάθε φορά από το πλήθος των δεδομένων. [33] Το λογισμικό αυτό περιέχει μία Διαδικτυακή Διεπαφή Χρήστη, η οποία αποτελεί το περιβάλλον στο οποίο εργάζεται ο χρήστης. Η Διεπαφή αυτή είναι προσβάσιμη από τους περισσότερους φυλλομετρητές. Η λειτουργία του λογισμικού βασίζεται στην χρήση servers στους οποίους αποθηκεύονται τα δεδομένα του Αρχείου Καταγραφής, τα μοντέλα, άλλα δεδομένα και άλλα αποτελέσματα, στην χρήση ερωτημάτων SQL (SQL Queries) για την πραγματοποίηση των υπολογισμών, αρχεία κώδικα για την εξαγωγή, μετασχηματισμό και φόρτωση δεδομένων (ETL Scripts). Συνοπτικά σε μία βάση δεδομένων ενός SQL server (SQL Server datatables database) αποθηκεύονται τα δεδομένα που αφορούν Γεγονότα και Υποθέσεις και άλλα δεδομένα που πρόκειται να χρησιμοποιηθούν στα αρχεία κώδικα για την εξαγωγή, μετασχηματισμό και φόρτωση δεδομένων (ETL Scripts), αυτά τα αρχεία κώδικα εκτελούνται σε μία βάση δεδομένων αρχείων κώδικα ενός SQL server (SQL Server scripting database), ενώ σε μία βάση μετα-δεδομένων ενός SQL server (SQL

Server metadata database) αποθηκεύονται τα projects που έχουν δημιουργηθεί, οι ρυθμίσεις του χρήστη κ.α. Σε έναν άλλο server, τον QPR ProcessAnalyzer Server, ο οποίος είναι στοιχείο cloud, δημιουργούνται τα απαραίτητα ερωτήματα SQL (SQL Queries) και αποτελεί ουσιαστικά το περιβάλλον όπου συντονίζονται οι επεξεργασίες και οι υπολογισμοί των δεδομένων τα οποία αποθηκεύονται στους 3 προαναφερθέντες servers. Το περιβάλλον στο οποίο αλληλεπιδρά ο χρήστης με το λογισμικό είναι η Διαδικτυακή Διεπαφή Χρήστη, όπου σε ένα φιλικό προς τον χρήστη περιβάλλον ο χρήστης μπορεί να εισάγει τα δεδομένα που επιθυμεί, κάποιο μοντέλο, να ορίσει φίλτρα για τα δεδομένα που εμφανίζονται, να κάνει διάφορες αναλύσεις δεδομένων κλπ. Για τις περιπτώσεις όπου τα δεδομένα εξάγονται από τοπικά (και όχι διαδικτυακά) συστήματα, χρησιμοποιείται επιπλέον ένας εκινητής αρχείων κώδικα (QPR ScriptLauncher) για την εξαγωγή δεδομένων και την αποθήκευσή τους στο cloud περιβάλλον του QPR ProcessAnalyzer. Τέλος προσφέρεται και μία εξωτερική υπηρεσία cloud, το Snowflake data cloud, η οποία όταν ενσωματώνεται με το λογισμικό QPR ProcessAnalyzer μπορεί να αποθηκεύει πίνακες δεδομένων, να διενεργεί υπολογισμούς και να εκτελεί κώδικες για την εξαγωγή, μετασχηματισμό και φόρτωση δεδομένων. [34]

Κεφάλαιο 4

Προβολή Δεδομένων και Ανάλυση

Αφετηρία της μελέτης είναι ένα Αρχείο Καταγραφής Γεγονότων σε μορφή Excel το οποίο περιέχει 89924 Γεγονότα που έχουν καταγραφεί από τον Ιούνιο του 2019 έως τον Μάιο του 2020 για μία Διαδικασία έγκρισης δανείων.

4.1 Το Αρχείο Καταγραφής Γεγονότων

Κάθε γραμμή του Φύλλου Excel είναι ένα γεγονός το οποίο έχει καταγραφεί αυτοματοποιημένα (δηλαδή μέσω κάποιου πληροφοριακού συστήματος) ή χειροκίνητα. Κάθε Γεγονός περιλαμβάνει υποχρεωτικά τα χαρακτηριστικά Case ID, Event Time και Start Time, δηλαδή το Αναγνωριστικό Υπόθεσης στην οποία ανήκει, τον Τύπο Γεγονότος και την ημερομηνία και ώρα έναρξης που συνέβη (ή ξεκίνησε να συμβαίνει). Κάθε γεγονός περιέχει επίσης το χαρακτηριστικό Automation το οποίο υποδηλώνει τον τρόπο με τον οποίο καταχωρήθηκε το γεγονός αυτό στο Αρχείο Καταγραφής (αυτόματα ή χειροκίνητα). Στην Εικόνα 4.1 φαίνεται ένα μέρος του αρχείου Excel το οποίο περιέχει το Αρχείο Καταγραφής Γεγονότων.

	A	B	C	D
1	Case ID	Event Type	Start Time	Automation
2	16613129	New Loan Application	2019-07-05 20:48:31.000	Automatic
3	16613129	Process Loan Application	2019-07-08 00:29:04.000	Manual
4	16613129	B2C First Approval	2019-07-10 21:03:06.000	Manual
5	16613129	Loan Rejected	2019-07-11 22:49:56.000	Manual
6	16613139	New Loan Application	2019-07-05 21:33:27.000	Manual
7	16613139	Process Loan Application	2019-07-05 22:30:06.000	Manual
8	16613139	B2C First Approval	2019-07-05 23:10:29.000	Manual
9	16613139	Loan Approved	2019-07-06 00:12:11.000	Manual
10	16613139	Manual Check for Valid Doc	2019-07-07 23:35:06.000	Manual
11	16613139	Process Loan to Customer	2019-07-08 00:38:31.000	Automatic
12	16613149	New Loan Application	2019-07-05 22:02:05.000	Manual
13	16613149	Process Loan Application	2019-07-05 22:36:16.000	Manual
14	16613149	Application Changed	2019-07-05 22:56:14.000	Manual
15	16613149	Process Loan Application	2019-07-05 23:37:35.000	Manual
16	16613149	B2C First Approval	2019-07-05 23:41:45.000	Manual
17	16613149	Loan Approved	2019-07-06 00:00:38.000	Automatic
18	16613149	Automated Check for Valid	2019-07-06 17:50:13.000	Automatic
19	16613149	Process Loan to Customer	2019-07-06 18:53:00.000	Manual
20	16613149	Automated Check for Valid	2019-07-06 19:14:28.000	Automatic
21	16613149	Process Loan to Customer	2019-07-06 19:14:39.000	Automatic
22	16613159	New Loan Application	2019-07-05 22:41:24.000	Automatic
23	16613159	Process Loan Application	2019-07-05 23:25:25.000	Manual
24	16613159	B2C First Approval	2019-07-05 23:33:39.000	Manual
25	16613159	Application Changed	2019-07-10 18:30:15.000	Manual

Εικόνα 4.1: Αρχείο Καταγραφής Γεγονότων

Το αναγνωριστικό Υπόθεσης είναι ένας οκταψήφιος κωδικός, με βάση τον οποίο κάθε Υπόθεση ορίζεται με μοναδικό τρόπο.

Οι Τύποι Γεγονότων είναι οι εξής:

- Αίτηση Νέου Δανείου - New Loan Application.
- Διεκπεραίωση Αίτησης Δανείου - Process Loan Application
- Απόρριψη Δανείου - Loan Rejected
- Έγκριση Δανείου - Loan Approved
- Χειροκίνητος Έλεγχος για την Εγκυρότητα των Εγγράφων - Manual Check for Valid Documentation
- Αυτοματοποιημένος Έλεγχος για την Εγκυρότητα των Εγγράφων - Automated Check for Valid Documentation
- Διαδικασία Απόδοσης Δανείου σε Λογαριασμό Πελάτη - Process Loan to Customer Account
- Τροποποίηση Αίτησης - Application Changed
- Επισύναψη Πρόσθετων Εγγράφων - Attach Additional Documents
- Πρώτη Έγκριση μεταξύ Επιχειρήσεων - B2B First Approval
- Δεύτερη Έγκριση μεταξύ Επιχειρήσεων - B2B Second Approval
- Πρώτη Έγκριση μεταξύ Επιχείρησης και Πελάτη- B2C First Approval
- Δεύτερη Έγκριση μεταξύ Επιχείρησης και Πελάτη- B2C Second Approval
- Εξωτερικός Σύμβουλος - External Consultant
- Ακύρωση Αίτησης - Application Cancelled
- Επικαιροποίηση Αίτησης - Application Updated
- Ο Πελάτης επισυνάπτει Πρόσθετα Έγγραφα - Customer Appends Documentation
- Ο Πελάτης ζητά Περαιτέρω Αξιολόγηση - Customer Asks for Further Assessment
- Περαιτέρω αξιολόγηση - Further Assessment

4.2 Εφαρμογή Μεθόδων Εξόρυξης Διεργασιών

4.2.1 Ανακάλυψη του Μοντέλου της Διεργασίας

Οι Τύποι Γεγονότων που παρουσιάζονται στην Ενότητα 4.1 στην πραγματικότητα είναι οι ονομασίες των Δραστηριοτήτων της Διεργασίας. Μία λογική σειρά βημάτων είναι η ακολουθία Αίτηση Νέου Δανείου → Διεκπεραίωση Αίτησης Δανείου → Έλεγχος για την Εγκυρότητα των Εγγράφων (Αυτοματοποιημένος ή Χειροκίνητος) → Έγκριση Δανείου. Βλέπουμε ωστόσο ότι υπάρχουν κι άλλα βήματα. Ενδεχομένως μετά τον Έλεγχο των Εγγράφων να ακολουθεί Επισύναψη Πρόσθετων Εγγράφων, να χρειαστεί η συμβολή Εξωτερικού Συμβούλου ή Έγκριση

μεταξύ Επιχείρησης και Πελάτη κ.ο.κ. Είναι λοιπόν χρήσιμο να έχουμε μία σαφή εικόνα του τρόπου με τον οποίο εκτελείται η Διεργασία που μελετάμε και για τον σκοπό αυτό θα χρησιμοποιήσουμε τεχνικές εξόρυξης διεργασιών.

Πριν περάσουμε στην εξαγωγή του Μοντέλου το οποίο περιγράφει την Διεργασία, είναι χρήσιμο να παρουσιάσουμε μία συνοπτική επισκόπηση του Αρχείου Καταγραφής Γεγονότων με την βοήθεια του εργαλείου ProM (βλ. Ενότητα 3.2). Όπως φαίνεται στις Εικόνες 4.2 και 4.3, το Αρχείο Καταγραφής Γεγονότων περιέχει 14032 Υποθέσεις (Process Instances) και 89924 Γεγονότα (Events), τα οποία ανήκουν σε 19 κατηγορίες ή αλλιώς Τύπους Γεγονότων, όπως αυτοί έχουν αναφερθεί στην Ενότητα 4.1. Οι τύποι Γεγονότων ουσιαστικά είναι οι Δραστηριότητες ή αλλιώς τα βήματα της Διεργασίας. Στην Εικόνα 4.2 φαίνεται μία λίστα με τους Τύπους Γεγονότων ταξινομημένους σε φθίνουσα σειρά με βάση των αριθμό εμφανίσεων στο Αρχείο Καταγραφής Γεγονότων. Οι τρεις συχνότεροι Τύποι Γεγονότων είναι η Διεκπεραίωση της Αίτησης Δανείου (Process Loan Application) με 15344 εμφανίσεις στο Αρχείο Καταγραφής Γεγονότων, η Αίτηση Νέου Δανείου (New Loan Application) με 14034 εμφανίσεις και η Πρώτη Έγκριση μεταξύ Επιχείρησης και Πελάτη (B2C First Approval) με 12167 εμφανίσεις. Από αυτό είναι φανερό ότι σε ορισμένες Υποθέσεις υπάρχει επανάληψη βημάτων, καθώς οι εμφανίσεις ορισμένων Δραστηριοτήτων υπερβαίνουν τον αριθμό Υποθέσεων του Αρχείου Καταγραφής. Στην Εικόνα 4.3 φαίνεται μία λίστα με τους Τύπους Γεγονότων που αποτελούν Γεγονότα Έναρξης και Λήξης της Διεργασίας, ταξινομημένους σε φθίνουσα σειρά με βάση τον αριθμό εμφανίσεων στις οποίες το κάθε Γεγονός αποτελεί Γεγονός έναρξης και λήξης αντίστοιχα. Ένα Γεγονός είναι Γεγονός Έναρξης μιας Υπόθεσης, αν δεν υπάρχει άλλο Γεγονός που να ανήκει στην ίδια Υπόθεση (δηλαδή να έχει την ίδια τιμή στο χαρακτηριστικό «Αγνωριστικό Υπόθεσης» (Case ID) και να έχει Ημερομηνία και Ώρα Έναρξης (Start Time) που να προηγείται από αυτήν του Γεγονότος Έναρξης, ενώ αντίστοιχα ένα Γεγονός είναι Γεγονός Λήξης μιας Υπόθεσης, αν δεν υπάρχει άλλο Γεγονός που να ανήκει στην ίδια Υπόθεση και να έχει Ημερομηνία και Ώρα Έναρξης (Start Time) που να έπεται από αυτήν του Γεγονότος Λήξης. Ουσιαστικά τα Γεγονότα Έναρξης και Λήξης μας δείχνουν με ποιο βήμα ξεκινάει η Διαδικασία και ποιο βήμα σημαίνει την λήξη της. Η Αίτηση Νέου Δανείου αποτελεί Γεγονός Έναρξης σε 13868 Υποθέσεις, ενώ η Διαδικασία Απόδοσης Δανείου σε Λογαριασμό Πελάτη (Process Loan to Customer Account) αποτελεί Γεγονός Λήξης σε 8179 Υποθέσεις.

Στο σημείο αυτό προχωράμε σε μία πρώτη προσέγγιση του Μοντέλου παρουσιάζοντας ένα Γράφημα Άμεσης Ακολουθίας το οποίο εξάγεται με βάση την συχνότητα εμφάνισης της ακολουθίας δύο Δραστηριοτήτων στο Αρχείο Καταγραφής Γεγονότων. Με αυτόν το τρόπο έχουμε ένα γράφημα με την συνηθέστερη συμπεριφορά του Αρχείου Καταγραφής και μία πρώτη εικόνα του Μοντέλου. Στην περίπτωση μας ορίζουμε ως τιμή κατωφλίου το 0.1, δηλαδή μας ενδιαφέρουν μόνο οι περιπτώσεις Δραστηριοτήτων που ακολουθούν η μία την άλλη τουλάχιστον 1404 φορές (10% επί των συνολικών Υποθέσεων) μέσα στο Αρχείο Καταγραφής Γεγονότων. Το γράφημα που προκύπτει φαίνεται στην Εικόνα 4.4.

Τα ορθογώνια πλαίσια του Γραφήματος Άμεσης Ακολουθίας της Εικόνας 4.4 απεικονίζουν τύπους Γεγονότων, δηλαδή βήματα της Διαδικασίας, ενώ τα βέλη απεικονίζουν την φορά των βημάτων. Μέσα στα ορθογώνια πλαίσια αναγράφεται ο τύπος του γεγονότος, ή αλλιώς το όνομα της Δραστηριότητας, ενώ πάνω στα βέλη αναγράφεται ο αριθμός των εμφανίσεων του κάθε βήματος. Διαπιστώνουμε πως το εναρκτήριο βήμα της Διεργασίας για

Log Summary

Total number of process instances: 14032
Total number of events: 89924

Event Name

Event classes defined by Event Name
All events

Total number of classes: 19

Class	Occurrences (absolute)	Occurrences (relative)
Process Loan Application	15344	17.063%
New Loan Application	14034	15.607%
B2C First Approval	12167	13.53%
Loan Approved	9530	10.598%
Process Loan to Customer Account	8816	9.804%
Application Changed	7195	8.001%
Manual Check for Valid Documentation	6730	7.484%
Loan Rejected	5391	5.995%
Automated Check for Valid Documentation	3678	4.09%
B2B First Approval	3468	3.857%
Attach Additional Documents	1103	1.227%
Customer Appends Documentation	859	0.955%
External Consultant	550	0.612%
Customer Asks for Further Assessment	299	0.333%
Further Assessment	220	0.245%
Application Updated	189	0.21%

Εικόνα 4.2: Συνοπτική Επισκόπηση του Αρχείου Καταγραφής στο περιβάλλον του ProM

Log Summary

Start events

Total number of classes: 3

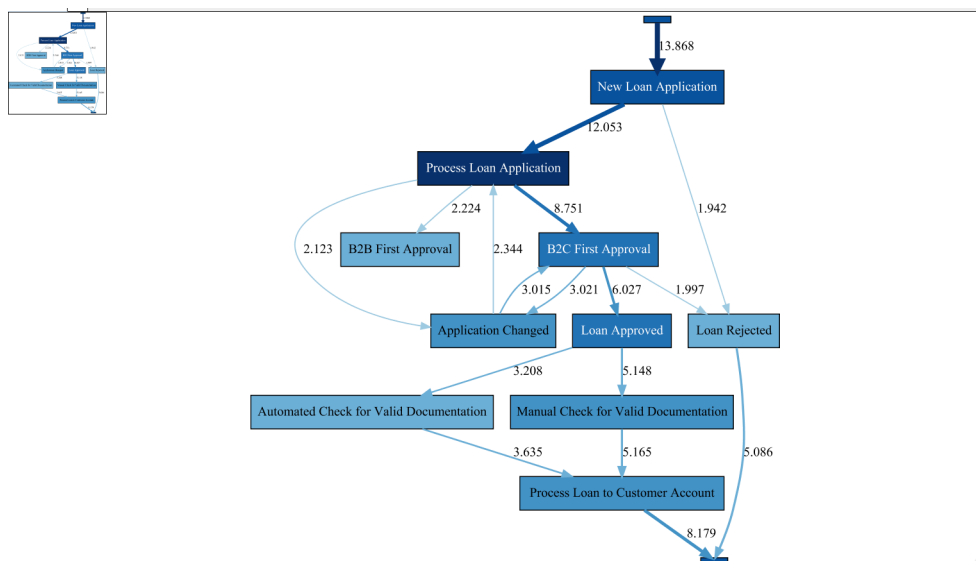
Class	Occurrences (absolute)	Occurrences (relative)
New Loan Application	13868	98.831%
Application Cancelled	163	1.162%
Loan Rejected	1	0.007%

End events

Total number of classes: 15

Class	Occurrences (absolute)	Occurrences (relative)
Process Loan to Customer Account	8179	58.288%
Loan Rejected	5086	36.246%
Loan Approved	527	3.756%
Manual Check for Valid Documentation	78	0.556%
Automated Check for Valid Documentation	43	0.305%
Attach Additional Documents	31	0.221%
Customer Asks for Further Assessment	25	0.178%
Customer Appends Documentation	19	0.135%
Application Changed	16	0.114%
B2C First Approval	14	0.1%
Further Assessment	9	0.064%
External Consultant	2	0.014%
B2B First Approval	1	0.007%
Process Loan Application	1	0.007%
New Loan Application	1	0.007%

Εικόνα 4.3: Συνοπτική Επισκόπηση του Αρχείου Καταγραφής στο περιβάλλον του ProM



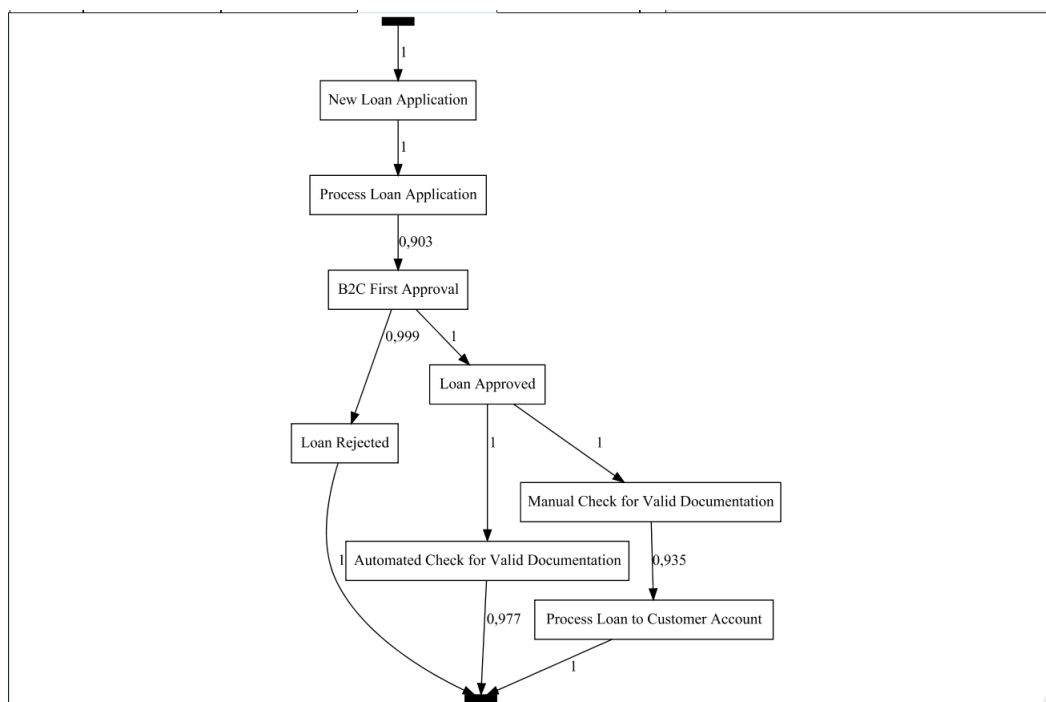
Εικόνα 4.4: Απεικόνιση του Μοντέλου ως Γράφημα Άμεσης Ακολουθίας

τις περισσότερες Υποθέσεις (13868) είναι η Αίτηση Νέου Δανείου (New Loan Application) και το τελικό βήμα αντίστοιχα είναι η Διαδικασία Απόδοσης Δανείου σε Λογαριασμό Πελάτη (Process Loan to Customer Account) (8179 Υποθέσεις), όπως άλλωστε επιβεβαιώνεται και από την Επισκόπηση του Αρχείου Καταγραφής Γεγονότων που έχει προηγηθεί. Η πιο συχνή ακολουθία βημάτων είναι η πραγματοποίηση της Διεκπεραίωσης της Αίτησης Δανείου (Process Loan Application) αμέσως μετά την Αίτηση Νέου Δανείου (New Loan Application).

Εφόσον έχει παρουσιαστεί μία πρώτη εικόνα του Μοντέλου, μπορούμε να προχωρήσουμε στην ανακάλυψη του Μοντέλου χρησιμοποιώντας τον Ευρετικό Αλγόριθμο (βλ. Ενότητα 2.1.1), ο οποίος υλοποιείται στο περιβάλλον του ProM. Σύμφωνα με τον Ευρετικό Αλγόριθμο, μία Δραστηριότητα B ακολουθεί άμεσα μία δραστηριότητα A, αν η τιμή της μετρικής $A \Rightarrow_w B$ υπερβαίνει μία τιμή που έχουμε θέσει ως τιμή κατωφλίου, στην περίπτωσή μας η τιμή αυτή είναι το 0.9, με τον τύπο της μετρικής να δίνεται από την σχέση: $A \Rightarrow_w B \equiv \left(\frac{|A>_w B| - |B>_w A|}{|A>_w B| + |B>_w A| + 1} \right)$. Επίσης θέτουμε ως τιμή κατωφλίου για τις μετρικές που σχετίζονται με τους βρόγχους μήκους 1 και 2 και πάλι το 0.9. Ως αποτέλεσμα εξάγεται το γράφημα της Εικόνας 4.5.

Στο Μοντέλο της Εικόνας 4.5 απεικονίζονται και οι 19 Δραστηριότητες της Διεργασίας, ενώ πάνω στα βέλη που τις συνδέουν φαίνεται η τιμή της μετρικής $A \Rightarrow_w B$ για κάθε ζεύγος Δραστηριοτήτων A και B. Αξίζει να προσέξουμε την περίπτωση των Δραστηριοτήτων B2C First Approval και Application Changed. Στο Γράφημα Άμεσης Ακολουθίας της Εικόνας 4.4 μπορούμε να δούμε ότι η Δραστηριότητα Application Changed εμφανίζεται αμέσως μετά την B2C First Approval σε 3021 Ίχνη και το αντίστροφο συμβαίνει σε 3015 Ίχνη. Συνεπώς αν συμβολίσουμε με AC την Δραστηριότητα Application Changed και με BFA την Δραστηριότητα B2C First Approval, τότε $AC \Rightarrow_w BFA = \left(\frac{|AC>_w BFA| - |BFA>_w AC|}{|AC>_w BFA| + |BFA>_w AC| + 1} \right) = \left(\frac{3021 - 3015}{3021 + 3015 + 1} \right) = 0,0009$, δηλαδή η τιμή της μετρικής βρίσκεται πολύ κοντά στο 0 και συνεπώς στο Γράφημα της Εικόνας 4.5 δεν υπάρχει βέλος που να συνδέει τις δύο δραστηριότητες. Ενδιαφέρον έχει επίσης η περίπτωση της Δραστηριότητας Attach Additional Documents, η οποία στο Γράφημα της Εικόνας 4.5 συνδέεται με την Δραστηριότητα Manual Check for Valid Documentation, παρόλο που όπως φαίνεται η τιμή της αντίστοιχης μετρικής είναι 0.014, δηλαδή δεν υπερβαίνει το

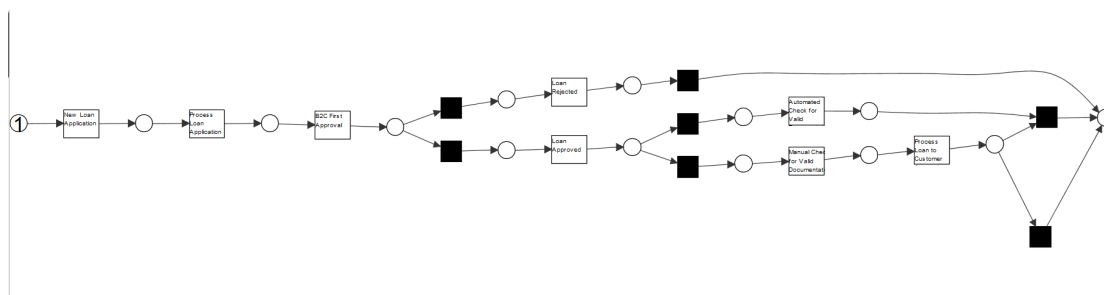
τουλάχιστον 2807 Ίχνη (20% των συνολικών Ίχνών).



Εικόνα 4.6: Ευρετικός Αλγόριθμος με πρόσθετους περιορισμούς

4.2.2 Έλεγχος Προσαρμογής

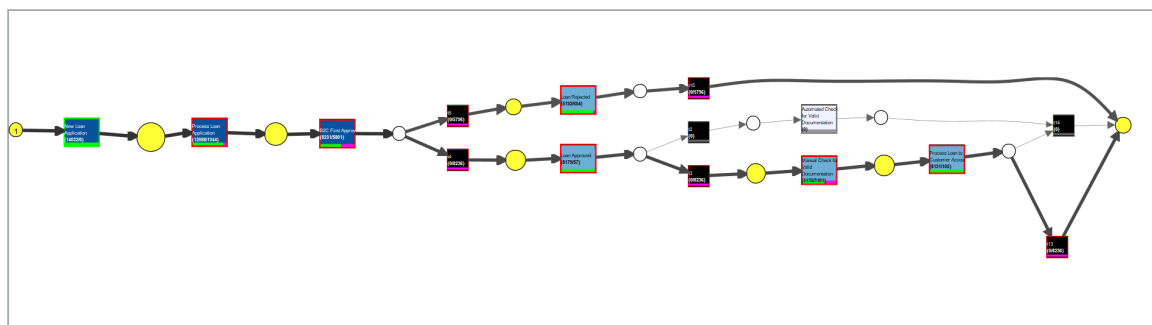
Σε αυτήν την Ενότητα πραγματοποιείται ο Έλεγχος Προσαρμογής μεταξύ ενός Μοντέλου που έχει σχεδιαστεί για την Διεργασία που εξετάζεται και του Αρχείου Καταγραφής το οποίο έχουμε στην διάθεσή μας. Για τον έλεγχο θα χρησιμοποιήσουμε το Μοντέλο M_1 , το οποίο φαίνεται στην Εικόνα 4.7 με την μορφή Δικτύου Petri και έχει εξαχθεί μέσω του Ευρετικού Αλγόριθμου με πρόσθετους περιορισμούς που εφαρμόστηκε στην Ενότητα 4.2.1 (βλ. Εικόνα 4.6).



Εικόνα 4.7: Δίκτυο Petri για το Μοντέλο M_1

Η ακολουθία των βημάτων της Διεργασίας σύμφωνα με το Μοντέλο M_1 είναι η εξής: New Loan Application → Process Loan Application → B2C First Approval και στο σημείο αυτό ακολουθεί είτε η Δραστηριότητα Loan Rejected, η οποία με την σειρά της οδηγεί σε λήξη της Διεργασίας, είτε η Loan Approved, μετά την οποία ακολουθεί είτε η Δραστηριότητα Automated Check for Valid Documentation με επόμενη Δραστηριότητα την λήξη της Διε-

γασίας, είτε η Manual Check for Valid Documentation με αμέσως επόμενη Δραστηριότητα την Process Loan to Customer Account και έπειτα λήξη της Διεργασίας. Με την βοήθεια του ProM θα ελέγξουμε την Προσαρμογή του M_1 με το Αρχείο Καταγραφής Γεγονότων που έχουμε στην διάθεσή μας εφαρμόζοντας την Μέθοδο της Ευθυγράμμισης (βλ. Ενότητα 2.2.1). Ορίζουμε την Τυπική Συνάρτηση Απόδοσης Κόστους Πιθανότητας με τρόπο ώστε να αποδίδει το ίδιο κόστος σε κάθε μετατόπιση και προχωράμε στον έλεγχο. Στην Εικόνα 4.8 φαίνεται το αποτέλεσμα του Ελέγχου Προσαρμογής. Οι θέσεις και μεταβάσεις του Δικτύου Petri που αναπαριστά το M_1 είναι χρωματισμένες με τρόπο ώστε να μας παρέχουν πληροφορίες για της Μετατοπίσεις στο Μοντέλο και στο Αρχείο Καταγραφής. Συγκεκριμένα όσο πιο βαθύ είναι το μπλε στις μεταβάσεις (ορθογώνια πλαίσια, Δραστηριότητες της Διεργασίας) τόσο πιο συχνά έχει εκτελεστεί αυτή η Δραστηριότητα στο Αρχείο Καταγραφής Γεγονότων. Παραδείγματος χάριν βλέπουμε ότι η Δραστηριότητα New Loan Application έχει εκτελεστεί περισσότερες φορές από την Δραστηριότητα Process Loan to Customer Account. Στο κάτω μέρος κάθε ορθογωνίου υπάρχει μία μπάρα χρωματισμένη με πράσινο ή/και ροζ. Το πράσινο χρώμα δείχνει το ποσοστό των Δραστηριοτήτων, οι εκτελέσεις των οποίων συμφωνούν στο Μοντέλο και στο Αρχείο Καταγραφής Γεγονότων, ενώ το ροζ χρώμα παριστάνει το ποσοστό των αποκλίσεων. Βλέπουμε ότι οι εκτελέσεις της Δραστηριότητας New Loan Application δεν παρουσιάζουν αποκλίσεις, ενώ για η Δραστηριότητα B2C First Approval αποκλίνει σε μεγάλο βαθμό. Οι θέσεις του Μοντέλου (κύκλοι) που είναι χρωματισμένες με κίτρινο χρώμα σημαίνουν ότι στο συγκεκριμένο σημείο συνέβη η εκτέλεση κάποιου Γεγονότος στο Αρχείο Καταγραφής, η οποία δεν μπορεί να εξηγηθεί με βάση το Μοντέλο. Παραδείγματος χάριν η θέση μετά την Δραστηριότητα New Loan Application έχει χρωματιστεί με κίτρινο χρώμα καθώς υπάρχουν Ίχνη του Αρχείου Καταγραφής στα οποία την εν λόγω Δραστηριότητα ακολουθεί η Δραστηριότητα Loan Rejected (βλ. Εικόνα 4.4), συμπεριφορά η οποία δεν εξηγείται με βάση το M_1 . Έτσι μπορούμε να έχουμε μία εικόνα των Μετατοπίσεων στο Μοντέλο και των Μετατοπίσεων στο Αρχείο Καταγραφής. Τα ορθογώνια πλαίσια μαύρου χρώματος που φαίνονται στο Δίκτυο Petri είναι μεταβάσεις οι οποίες συμβολίζουν την έννοια της διάζευξης στο γράφημα και δεν αντιστοιχούν σε κάποια Δραστηριότητα.



Εικόνα 4.8: Έλεγχος Προσαρμογής μεταξύ του Αρχείου Καταγραφής Γεγονότων και του Μοντέλου M_1

Ανοίγοντας το παράθυρο "Inspector" στο περιβάλλον του ProM μπορούμε να έχουμε μία πιο αναλυτική παρουσίαση των αποτελεσμάτων του Ελέγχου Προσαρμογής. Στην Εικόνα 4.9 φαίνονται κάποια στοιχεία του Ελέγχου. Συγκεκριμένα βλέπουμε ότι υπάρχουν συνολικά 14032 Ίχνη του Αρχείου Καταγραφής Γεγονότων τα οποία εξετάστηκαν, 60988 Σύγχρονες

Μετατοπίσεις (δηλαδή δεν υπήρχε απόκλιση μεταξύ του M_1 και του Αρχείου Καταγραφής), 47912 Μετατοπίσεις στο Μοντέλο και 28936 Μετατοπίσεις στο Αρχείο Καταγραφής. Σημειώνεται ότι το άθροισμα των Σύγχρονων Μετατοπίσεων και των Μετατοπίσεων στο Αρχείο Καταγραφής ισούται με το σύνολο των Γεγονότων του Αρχείου Καταγραφής, αφού κάθε Γεγονός του Αρχείου Καταγραφής είτε θα συμφωνεί με το Μοντέλο είτε θα εκτελείται με τρόπο που δεν ερμηνεύεται από το Μοντέλο. Επίσης βλέπουμε ότι το επίπεδο προσαρμογής είναι 75,32%.

Property	Value
#Moves log model	60086
#Moves model only	47912
#Moves log only	28936
#Traces	14032
#Infeasible Traces	0

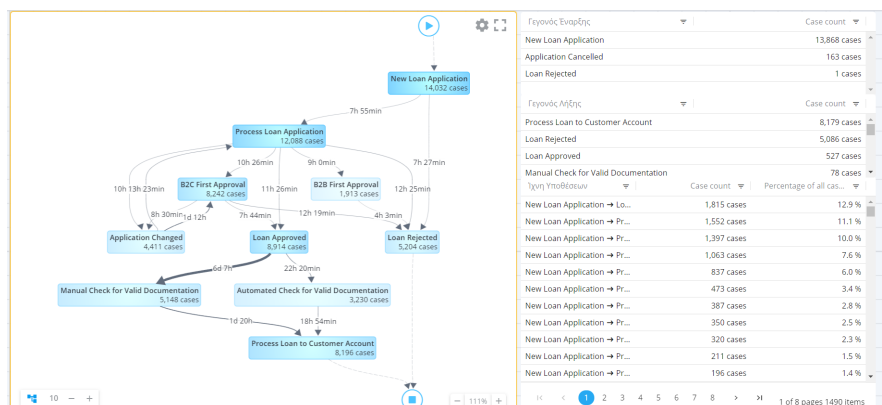
Property	Value
Calculation Time (ms)	0.551449258369422
Min Fitness Cost	2.508980270963554
Max Move-Log Cost	6.4084948688711725
Minn Status	10.374632321730837
Trace Fitness	0.752366567532538
Move Model Fitness	0.328689923983276
Move Log Fitness	0.752366567532538
Max Fitness Cost	19.408494868871173
Trace Length	0.4084948688711725

Εικόνα 4.9: Έλεγχος Προσαρμογής μεταξύ του Αρχείου Καταγραφής Γεγονότων και του Μοντέλου M_1 (Αναλυτικά Αποτελέσματα)

4.3 Εξόρυξη Διεργασιών σε Επιχειρησιακά Περιβάλλοντα

Στις Ενότητες 4.1, 4.2.1 και 4.2.2 παρουσιάστηκε η εφαρμογή ορισμένων μεθόδων σε ένα Αρχείο Καταγραφής Γεγονότων με την βοήθεια του εργαλείου ProM. Καθώς η Εξόρυξη Διεργασιών χρησιμοποιείται σε μεγάλο βαθμό από επιχειρήσεις (βλ. Κεφάλαιο 1), έχουν αναπτυχθεί διάφορα περιβάλλοντα τα οποία χρησιμοποιούν μεθόδους Εξόρυξης Διεργασιών με σκοπό την χρήση τους από επιχειρήσεις οι οποίες θέλουν να έχουν μία καλύτερη εικόνα των Διεργασιών που εκτελούνται στα πλαίσιά τους. Ένα τέτοιο εργαλείο είναι και το QPR Process Analyzer (βλ. Ενότητα 3.3), με την βοήθεια του οποίου θα παρουσιάσουμε στην συνέχεια ορισμένα αποτελέσματα που αφορούν το Αρχείο Καταγραφής Γεγονότων που παρουσιάζεται στην Ενότητα 4.1. Πέρα από την ανακάλυψη του Μοντέλου, με την βοήθεια του QPR Process Analyzer είμαστε σε θέση να παρουσιάσουμε ορισμένα στατιστικά του Αρχείου Καταγραφής Γεγονότων.

Στην Εικόνα 4.10 φαίνεται μία αναφορά που παράγεται στο εργαλείο QPR Process Analyzer, η οποία περιέχει πληροφορίες σχετικά με τα Γεγονότα Έναρξης και Λήξης της Διαδικασίας, καθώς και τα διαφορετικά Ίχνη των Υποθέσεων. Αριστερά φαίνεται το Διάγραμμα Ροής του Μοντέλου και Δεξιά τρεις Πίνακες οι οποίοι δείχνουν τα Γεγονότα τα οποία αποτελούν Γεγονός Έναρξης Υποθέσεων και τον αντίστοιχο αριθμό Υποθέσεων στις οποίες συμβαίνει αυτό (Πίνακας "Γεγονός Έναρξης"), τα Γεγονότα τα οποία αποτελούν Γεγονός Λήξης Υποθέσεων και τον αντίστοιχο αριθμό Υποθέσεων στις οποίες συμβαίνει αυτό (Πίνακας "Γεγονός Λήξης") και τα Ίχνη που υπάρχουν στο Αρχείο Καταγραφής και τον αντίστοιχο αριθμό Υποθέσεων αλλά και το ποσοστό αυτών (Πίνακας "Ίχνη Υποθέσεων"). Όλοι οι Πίνακες είναι ταξινομημένοι



Εικόνα 4.10: Αναφορά στο QPR Process Analyzer σχετικά με τα Γεγονότα Έναρξης και Λήξης και τα Ίχνη της Διαδικασίας

με βάση την συχνότητα σε φθίνουσα σειρά.

Το Ίχνος μιας Υπόθεσης προκύπτει ταξινομώντας τα Γεγονότα μιας Υπόθεσης (δηλαδή Γεγονότα με ίδια τιμή στο Case ID) με χρονολογική σειρά (δηλαδή από αυτό με την πιο παλιά ημερομηνία Έναρξης Start Time προς αυτό με την πιο πρόσφατη ημερομηνία Έναρξης Start Time).

Παρατηρούμε ότι το Ίχνος με την πιο συχνή εμφάνιση είναι αυτό που έχει ως αφετηρία το Γεγονός New Loan Application και ως πέρας το Γεγονός Loan Rejected, δηλαδή η απόρριψη του Δανείου. Ταυτόχρονα παρατηρούμε ότι οι Υποθέσεις στις οποίες τελικά το ποσό του Δανείου αποδόθηκε στον Λογαριασμό του Πελάτη (Δραστηριότητα Process Loan to Customer Account) είναι περισσότερες από τις περιπτώσεις στις οποίες τελικά το Δάνειο απορρίφθηκε (Δραστηριότητα Loan Rejected). Αυτό μας οδηγεί στο συμπέρασμα ότι τα περισσότερα Δάνεια στην πραγματικότητα εγκρίνονται, αλλά η διαδικασία μέχρι την Έγκριση μπορεί να περιλαμβάνει πολλά διαφορετικά βήματα, τα οποία έχουν ως αποτέλεσμα την ύπαρξη διαφορετικών Ιχνών, ενώ η Διαδικασία μέχρι την Απόρριψη συνήθως γίνεται αμέσως μετά την νέα Αίτηση. Επιπλέον από τον Πίνακα "Ίχνη Υποθέσεων" βλέπουμε μπορούν να εξαχθούν συμπεράσματα σχετικά με τον αριθμό των Ιχνών που θα κρατήσουμε στο Διάγραμμα Ροής. Θέτοντας ως κατώφλι το 1.5% βλέπουμε ότι τα Ίχνη που εμφανίζονται το κάθε ένα σε Υποθέσεις που καλύπτουν παραπάνω από το 1.5% του μοντέλου είναι τα 10 πρώτα του Πίνακα.

4.3.1 Το Αρχείο Καταγραφής Υποθέσεων

Είδαμε ότι το πιο συχνό Ίχνος της Διαδικασίας είναι το Ίχνος από το New Loan Application στο Loan Rejected, δηλαδή η απόρριψη του Δανείου αμέσως μετά την Υποβολή της Αίτησης. Στο σημείο αυτό θα ήταν ιδιαίτερα χρήσιμο να εμπλουτιστούν τα δεδομένα με ένα Αρχείο Καταγραφής Υποθέσεων, ώστε να είναι στην διάθεσή μας χαρακτηριστικά Υποθέσεων τα οποία καθιστούν δυνατή την εξαγωγή περισσότερων συμπερασμάτων.

Το Αρχείο Καταγραφής Υποθέσεων είναι επίσης σε μορφή Excel και ακολουθεί την ίδια λογική με το Αρχείο Καταγραφής Γεγονότων, δηλαδή κάθε γραμμή του Αρχείου είναι μία διαφορετική Υπόθεση, η οποία περιέχει οπωσδήποτε το χαρακτηριστικό Case Id και πέρα από αυτό ορισμένα χαρακτηριστικά που αφορούν συνολικά την Υπόθεση. Σε αντίθεση με

τα Γεγονότα του Αρχείου Καταγραφής Γεγονότων, οι Υποθέσεις του Αρχείου Καταγραφής Υποθέσεων δεν περιέχουν κατ' ανάγκη χαρακτηριστικό που να δηλώνει χρονική σήμανση.

Είναι σαφές ότι τα δύο αρχεία πρέπει να αφορούν τις ίδιες Υποθέσεις, δηλαδή να περιέχουν τα ίδια Case ID. Κάθε Υπόθεση του Αρχείου με το οποίο θα εμπλουτιστεί το Μοντέλο περιλαμβάνει τα εξής χαρακτηριστικά:

- Αναγνωριστικό Υπόθεσης - Case ID
- Επιχειρηματική Μονάδα - Business Unit
- Κόστος - Cost
- Λόγος απόρριψης του Δανείου - Loan rejection reason
- Κατάσταση του Δανείου - Loan Status
- Τοποθεσία - Location
- Μήνας - Month
- Προϊόν - Product
- SLA²
- Χρήστης - User

Στην Εικόνα 4.11 παρουσιάζεται ένα μέρος του αρχείου Excel το οποίο περιέχει το Αρχείο Καταγραφής Υποθέσεων. Κάθε ένα από τα χαρακτηριστικά αυτά παρέχει πληροφορίες που αφορούν συνολικά την κάθε Αίτηση και όχι την εκτέλεση κάποιας Δραστηριότητας μεμονωμένα.

	A	B	C	D	E	F	G	H	I	J
1	Case ID	Business Unit	Cost	Loan rejection reason	Loan Status	Location	Month	Product	SLA	User
2	16612500	Merchant	3312	N/A	Loan approved and processed	Hong Kong	2019/10	Boat/ship loan (15%)	1.In Target	Alexander Serrano
3	16612501	Bank branch	11200	Not enough collateral	Loan rejected	Tyler	2019/06	Consumer loan (ZZ1)	2.Out of Target	Antwan Boyle
4	16612503	Merchant	544	N/A	Loan approved and processed	Hong Kong	2019/06	Consumer loan (XX1)	2.Out of Target	Angeline Yu
5	16612508	Merchant	2704	N/A	Loan application in progress	Hong Kong	2019/06	Boat/ship loan (15%)	1.In Target	Ronan Benson
6	16612510	Bank branch	8000	Poor credit score	Loan rejected	Frisco	2019/06	Consumer loan (ZZ1)	1.In Target	Matthias Freeman
7	16612511	Bank branch	11200	N/A	Loan approved and processed	Sevilla	2019/06	Consumer loan (ZZ1)	2.Out of Target	Kierra Fields
8	16612512	Bank branch	5120	N/A	Loan approved and processed	Toledo	2019/06	Consumer loan (ZZ1)	2.Out of Target	Kamora Shields
9	16612516	Merchant	2877	Poor credit score	Loan rejected	Hong Kong	2019/06	Consumer loan (XX1)	2.Out of Target	Sarah Valdez
10	16612517	Merchant	640	N/A	Loan approved and processed	Colorado Springs	2019/06	Immediate loan (XX%)	2.Out of Target	Sarah Valdez
11	16612520	Bank branch	2400	Other/unclassified	Loan rejected	Kishinev	2019/06	Consumer loan (ZZ1)	2.Out of Target	Bruce Ali
12	16612521	Bank branch	9600	Poor credit score	Loan rejected	Orange	2019/06	Consumer loan (XX1)	2.Out of Target	Justine Cantu
13	16612525	Bank branch	8000	Too risky industry	Loan rejected	Denton	2019/06	Consumer loan (XX1)	1.In Target	Demarcus Escobar
14	16612529	Merchant	8000	N/A	Loan approved and processed	Glasgow	2019/06	Car loan used (12%)	2.Out of Target	Liliana Sawyer
15	16612533	Bank branch	4320	N/A	Loan approved and processed	Essen	2019/06	Consumer loan (XX1)	2.Out of Target	Willow Barry
16	16612535	Bank branch	12800	Not enough cashflow	Loan rejected	Rotterdam	2019/06	Consumer loan (ZZ1)	2.Out of Target	Aidan Suarez
17	16612541	Bank branch	7200	N/A	Loan approved and processed	Sioux Falls	2019/06	Consumer loan (ZZ1)	2.Out of Target	Mason Hamilton
18	16612544	Bank branch	16000	Poor credit score	Loan rejected	Huntsville	2019/06	Consumer loan (ZZ1)	2.Out of Target	Noe Moody
19	16612546	Bank branch	6400	N/A	Loan approved and processed	Poznan	2019/06	Consumer loan (ZZ1)	2.Out of Target	Tyson Hicks
20	16612549	Merchant	5600	N/A	Loan approved and processed	Reno	2019/06	Car loan used (12%)	2.Out of Target	Serenity Graves
21	16612553	Bank branch	8000	Poor credit score	Loan rejected	Helsinki	2019/06	Consumer loan (ZZ1)	2.Out of Target	Vanessa Townsend
22	16612554	Bank branch	6400	N/A	Loan approved and processed	Marseille	2019/06	Consumer loan (XX1)	2.Out of Target	Eliza Graves
23	16612556	Merchant	4000	N/A	Loan approved and processed	Bremen	2019/06	Car loan used (12%)	2.Out of Target	Danica Mcpherson

Εικόνα 4.11: Αρχείο Καταγραφής Υποθέσεων

Οι τιμές του χαρακτηριστικού Επιχειρηματική Μονάδα αναφέρονται στην Μονάδα της Τράπεζας στην οποία έλαβε χώρα κάποια Υπόθεση και (Business Unit) είναι οι εξής:

² service-level agreement. Μια συμφωνία επιπέδου υπηρεσιών (SLA) είναι μια συμφωνία μεταξύ ενός παρόχου υπηρεσιών και ενός πελάτη. Συγκεκριμένες πτυχές της υπηρεσίας - ποιότητα, διαθεσιμότητα, ευθύνες, χρόνος - συμφωνούνται μεταξύ του παρόχου υπηρεσιών και του χρήστη της υπηρεσίας.

- Bank Branch - Υποκατάστημα Τράπεζας
- Call Center - Τηλεφωνικό Κέντρο
- Merchant - Εμπορικό
- Online Banking - Ηλεκτρονικές Τραπεζικές Υπηρεσίες

Οι τιμές του χαρακτηριστικού Κόστος (Cost) αναφέρονται στο ύψος του Δανείου. Οι τιμές του χαρακτηριστικού Λόγος Απόρριψης Δανείου (Loan Rejection Reason) είναι οι εξής:

- Incomplete Paperwork - Ελλιπή Έγγραφα
- Not enough collateral - Μη επαρκείς εξασφαλίσεις
- Not enough cashflow - Μη επαρκής ταμειακή ροή
- Poor credit score - Κακή πιστωτική βαθμολογία
- Too high debt distribution - Πολύ υψηλή κατανομή χρέους
- Too risky industry - Πολύ επικίνδυνος κλάδος
- Other/unclassified - Άλλο/άγνωστο
- N/A - Δεν υπάρχει

Οι τιμές του χαρακτηριστικού Κατάσταση Δανείου (Loan Status) αναφέρονται στην κατάσταση στην οποία βρίσκεται κάποια Υπόθεση την παρούσα χρονική στιγμή και είναι οι εξής:

- Loan application in progress - Αίτηση Δανείου σε εξέλιξη³
- Loan approved and processed - Έγκριση και Επεξεργασία Δανείου
- Loan rejected - Απόρριψη Δανείου

Οι τιμές του χαρακτηριστικού Τοποθεσία (Location) αναφέρονται στην τοποθεσία του υποκαταστήματος στο οποίο διεκπεραιώθηκε κάθε Υπόθεση.

Οι τιμές του χαρακτηριστικού Μήνας (Month) αναφέρονται στον μήνα κατά τον οποίο έγινε η Αίτηση και περιλαμβάνουν τους μήνες 6/2019 έως και 5/2020.

Οι τιμές του χαρακτηριστικού Προϊόν (Product) αναφέρονται στα διαφορετικά τραπεζικά προϊόντα της τράπεζας που αφορούν δάνεια και είναι οι εξής⁴:

- Boat/ship loan (15%) - Δάνειο για Βάρκα/Πλοίο (15%)
- Car loan new (8%) - Δάνειο για Νέο Αυτοκίνητο (8%)
- Car loan used (12%) - Δάνειο για Μεταχειρισμένο Αυτοκίνητο (12%)

³Ο χαρακτηρισμός "Σε εξέλιξη αφορά τον χαρακτηρισμό που είχε δοθεί σε κάθε αίτηση την στιγμή όπου αποκτήθηκε το αρχείο. Εφόσον το αρχείο αυτό δεν ανανεώνεται, η τιμή αυτή δεν αντικατοπτρίζει την πραγματικότητα με βάση την σημερινή ημερομηνία."

⁴Η λεπτομερής ανάλυση του κάθε προϊόντος (τι είναι και σε ποιες περιπτώσεις χορηγείται) δεν εμπίπτει στους σκοπούς της παρούσας εργασίας. Δεχόμαστε ότι υπάρχουν αυτά τα διαφορετικά προϊόντα όπως αυτά έχουν ονομαστεί από την τράπεζα στην οποία ανήκει το Αρχείο Καταγραφής

- Consumer loan (VIP) - Καταναλωτικό Δάνειο (VIP)
- Consumer loan (VIP2) - Καταναλωτικό Δάνειο (VIP2)
- Consumer loan (without background check) - Καταναλωτικό Δάνειο (χωρίς έλεγχο του υπόβαθρου)
- Consumer loan (XX1) - Καταναλωτικό Δάνειο (XX1)
- Consumer loan (XX3) - Καταναλωτικό Δάνειο (XX3)
- Consumer loan (XZ1) - Καταναλωτικό Δάνειο (XZ1)
- Consumer loan (XZ2) - Καταναλωτικό Δάνειο (XZ2)
- Consumer loan (ZZ1) - Καταναλωτικό Δάνειο (ZZ1)
- Consumer loan (ZZ2) - Καταναλωτικό Δάνειο (ZZ2)
- Immediate loan (XX%) - Άμεσο Δάνειο (XX%)
- Other consumer (high rate) - Άλλο Καταναλωτικό Δάνειο (υψηλό ποσοστό)
- Personal loan (2%) - Προσωπικό Δάνειο (2%)
- Personal loan (5%) - Προσωπικό Δάνειο (5%)
- Personal loan education (15%) - Προσωπικό Δάνειο για Εκπαίδευση (15%)

Οι τιμές του χαρακτηριστικού SLA παρέχουν πληροφορία για το αν η διεκπεραίωση της κάθε Υπόθεσης ανταπεξήλθε στους όρους της Συμφωνίας SLA και είναι οι εξής:

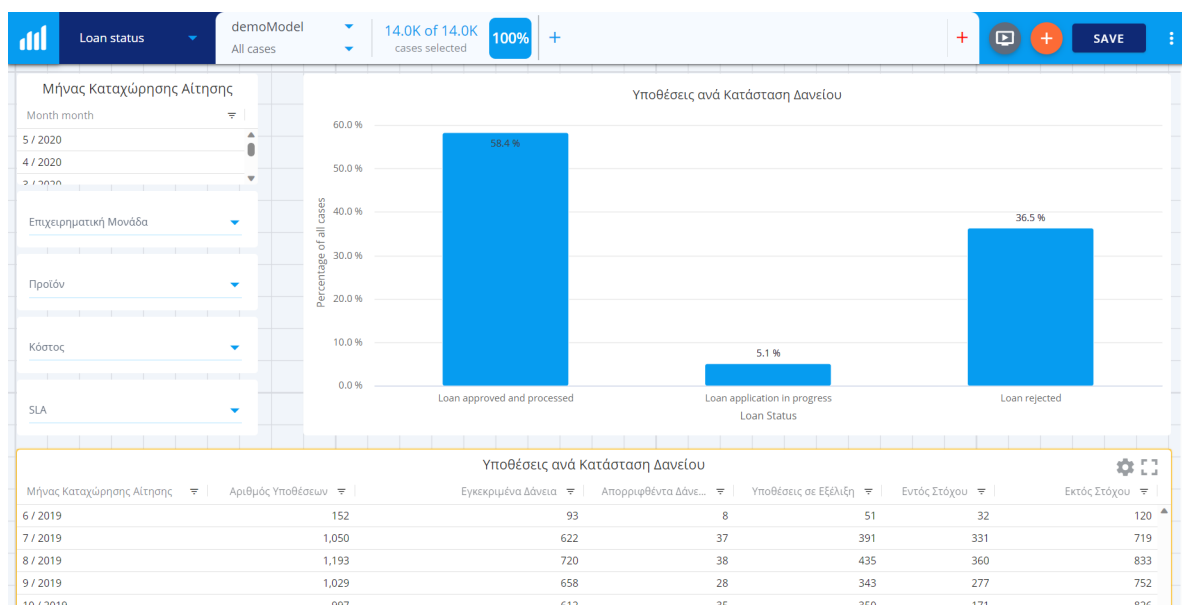
- In Target - Εντός Στόχου
- Out of Target - Εκτός Στόχου

Οι τιμές του χαρακτηριστικού Χρήστη (User) αναφέρονται στον Υπάλληλο που είναι υπεύθυνος για την κάθε Υπόθεση.

4.3.2 Αναφορές για τα Χαρακτηριστικά των Υποθέσεων

Στο σημείο αυτό είμαστε σε θέση να εξάγουμε ορισμένα συμπεράσματα για τον τρόπο με τον οποίο διεκπεραιώνεται η Διεργασία.

Στην Εικόνα 4.12 φαίνεται η αναφορά του QPR Process Analyzer για τις Υποθέσεις ανά Κατάσταση Δανείου. Το ραβδόγραμμα της αναφοράς δείχνει τα ποσοστά επί των συνολικών Υποθέσεων για κάθε μία από τις τιμές του χαρακτηριστικού Loan Status (Κατάσταση Δανείου). Ο Πίνακας που ακολουθεί δείχνει τους ακριβείς αριθμούς Υποθέσεων ανά μήνα καταχώρησης των Αιτήσεων για κάθε μία από τις κατηγορίες «Αριθμός Υποθέσεων», όπου φαίνεται ο συνολικός αριθμός Αιτήσεων που καταχωρήθηκαν κάθε μήνα, «Εγκεκριμένα Δάνεια», όπου φαίνεται ο αριθμός Δανείων με τιμή Loan approved and Processed στο χαρακτηριστικό Loan Status, «Απορριφθέντα Δάνεια», όπου φαίνεται ο αριθμός Δανείων με τιμή Loan rejected στο



Εικόνα 4.12: Αναφορά με βάση την Κατάσταση του Δανείου

χαρακτηριστικό Loan Status, «Υποθέσεις σε Εξέλιξη», όπου φαίνεται ο αριθμός Δανείων με τιμή Loan application in progress στο χαρακτηριστικό Loan Status, «Εντός Στόχου», όπου φαίνεται ο αριθμός Δανείων με τιμή In Target στο χαρακτηριστικό SLA και «Εκτός Στόχου» όπου φαίνεται ο αριθμός Δανείων με τιμή Out of Target στο χαρακτηριστικό SLA. Αριστερά φαίνονται τα φίλτρα «Μήνας Καταχώρησης Αίτησης», «Επιχειρηματική Μονάδα», «Προϊόν», «Κόστος» και «SLA».

Αρχικά το 58.4% των Δανείων που καταγράφηκαν έχουν εγκριθεί, το 36.5% έχουν απορριφθεί, ενώ το 5.1% των δανείων βρίσκονται ακόμα σε εξέλιξη. Από τον Πίνακα της αναφοράς που φαίνεται καλύτερα στην Εικόνα 4.13 βλέπουμε ότι οι περισσότερες αιτήσεις Δανείων καταχωρήθηκαν τον Δεκέμβριο του 2019 (1597 Αιτήσεις), με τις περισσότερες από αυτές να έχουν λήξει αισίως με έγκριση του Δανείου (1109 Υποθέσεις), αλλά και την πλειοψηφία να είναι Εκτός Στόχου SLA (1241 Υποθέσεις)

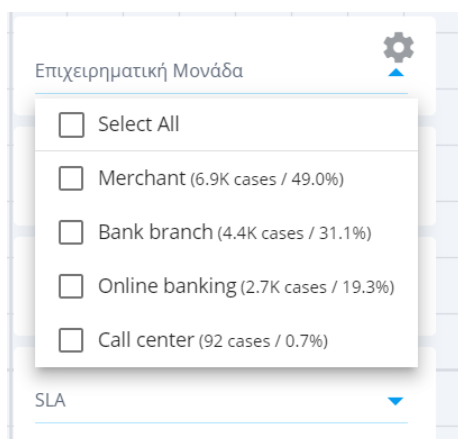
Με την βοήθεια των φίλτρων μπορεί κανείς να επικεντρωθεί σε συγκεκριμένες περιπτώσεις. Στο φίλτρο «Μήνας Καταχώρησης Αίτησης» μπορεί κανείς να επιλέξει τον ή τους μήνες που τον ενδιαφέρουν, ενώ στα υπόλοιπα φίλτρα εμφανίζεται μία λίστα επιλογών επιλέγοντας το βέλος που φαίνεται δίπλα στο όνομα κάθε φίλτρου (βλ. Εικόνα 4.14). Επιλέγοντας τα φίλτρα που θέλουμε αναπροσαρμόζονται το Γράφημα και ο Πίνακας της Αναφοράς και πλέον εμφανίζονται δεδομένα που ανταποκρίνονται στα κριτήρια που τέθηκαν. Κάθε φορά μπορούν να είναι επιλεγμένα πάνω από 1 φίλτρα.

Στην Εικόνα 4.15 φαίνεται η αναφορά των Υποθέσεων ανά Κατάσταση Δανείου έχοντας φιλτράρει τα δεδομένα με βάση τον μήνα καταχώρησης και την Επιχειρηματική Μονάδα. Συγκεκριμένα μας ενδιαφέρουν οι Αιτήσεις που καταχωρήθηκαν τους μήνες του 2020 από την Επιχειρηματική Μονάδα Merchant. Στην περίπτωση αυτή όπως φαίνεται οι περισσότερες Υποθέσεις είναι σε εξέλιξη.

Στην συνέχεια εξετάζουμε πιο αναλυτικά τις Υποθέσεις που αφορούν Δάνεια που έχουν απορριφθεί.

Μήνας Καταχώρησης Αίτησης	Αριθμός Υποθέσεων	Υποθέσεις ανά Κατάσταση Δανείου					Εκτός Στόχου
		Εγκριμένα Δάνεια	Απορριφθέντα Δάνεια	Υποθέσεις σε Εξέλιξη	Εντός Στόχου	Εκτός Στόχου	
6 / 2019	152	93	8	51	32	120	
7 / 2019	1,050	622	37	391	331	719	
8 / 2019	1,193	720	38	435	360	833	
9 / 2019	1,029	658	28	343	277	752	
10 / 2019	997	612	35	350	171	826	
11 / 2019	1,413	944	30	439	317	1,096	
12 / 2019	1,597	1,109	46	442	356	1,241	
1 / 2020	1,579	1,038	46	495	911	668	
2 / 2020	1,146	727	33	386	398	748	
3 / 2020	1,775	879	83	813	729	1,046	
4 / 2020	1,723	735	195	793	1,164	559	
5 / 2020	378	59	139	180	198	180	

Εικόνα 4.13: Πίνακας της Αναφοράς με βάση την Κατάσταση του Δανείου



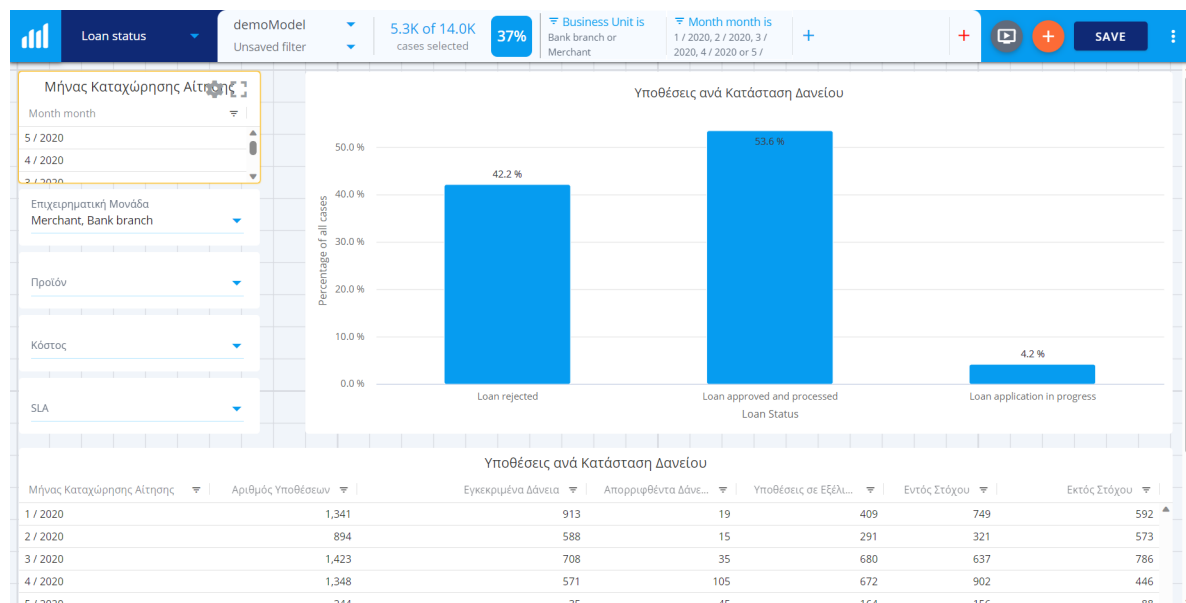
Εικόνα 4.14: Παράδειγμα επιλογής φίλτρου

Στην Εικόνα 4.16 φαίνεται η αναφορά για τις Υποθέσεις που αφορούν δάνεια που έχουν απορριφθεί. Το διάγραμμα δείχνει τα ποσοστά επί των απορριφθέντων δανείων για κάθε έναν από τους λόγους απόρριψης. Ο πίνακας της αναφοράς (βλ. Εικόνα 4.17) δείχνει τα Ίχνη του Αρχείου Καταγραφής, τα οποία αφορούν Υποθέσεις με Δάνεια που έχουν απορριφθεί, τον αριθμό Υποθέσεων με το κάθε Ίχνος και το αντίστοιχο ποσοστό επί των συνολικών απορριφθέντων δανείων. Όπως και στην προηγούμενη αναφορά υπάρχουν τα φίλτρα «Μήνας Καταχώρησης Αίτησης», «Επιχειρηματική Μονάδα», «Προϊόν», «Κόστος» και «SLA».

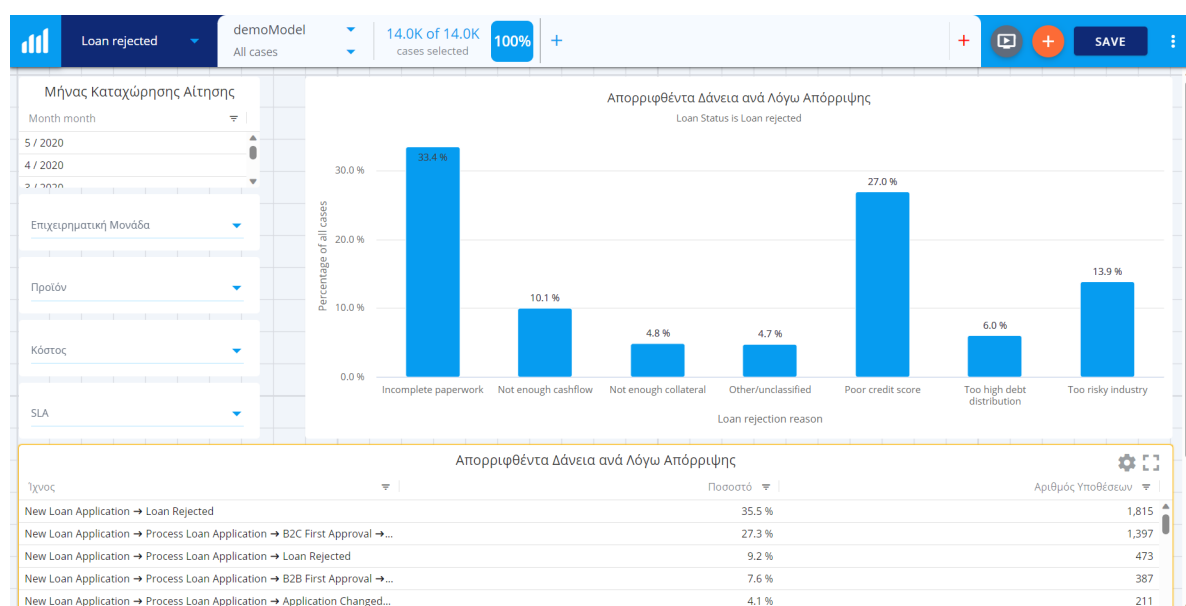
Το 33.4% των δανείων που απορρίπτονται, απορρίπτονται λόγω ελλιπών εγγράφων και αμέσως μετά ακολουθεί με 27% η κακή πιστωτική βαθμολογία. Το πιο σύνθηρες Ίχνος για Υποθέσεις με δάνεια που έχουν απορριφθεί είναι το < New Loan Application, Loan Rejected > και ακολουθεί το Ίχνος < New Loan Application, Process Loan Application, B2C First Approval, Loan Rejected >. Επιλέγοντας την πρώτη γραμμή του Πίνακα της αναφοράς τα δεδομένα φιλτράρονται ώστε να ανταποκρίνονται στις Υποθέσεις με το επιλεγμένο Ίχνος.

Στη συνέχεια περνάμε στην αναφορά που αφορά την διάρκεια των Υποθέσεων.

4.3.2 Αναφορές για τα Χαρακτηριστικά των Υποθέσεων



Εικόνα 4.15: Η Αναφορά έχοντας φιλτράρει τα δεδομένα με βάση τον Μήνα Καταχώρησης και Την Επιχειρηματική Μονάδα



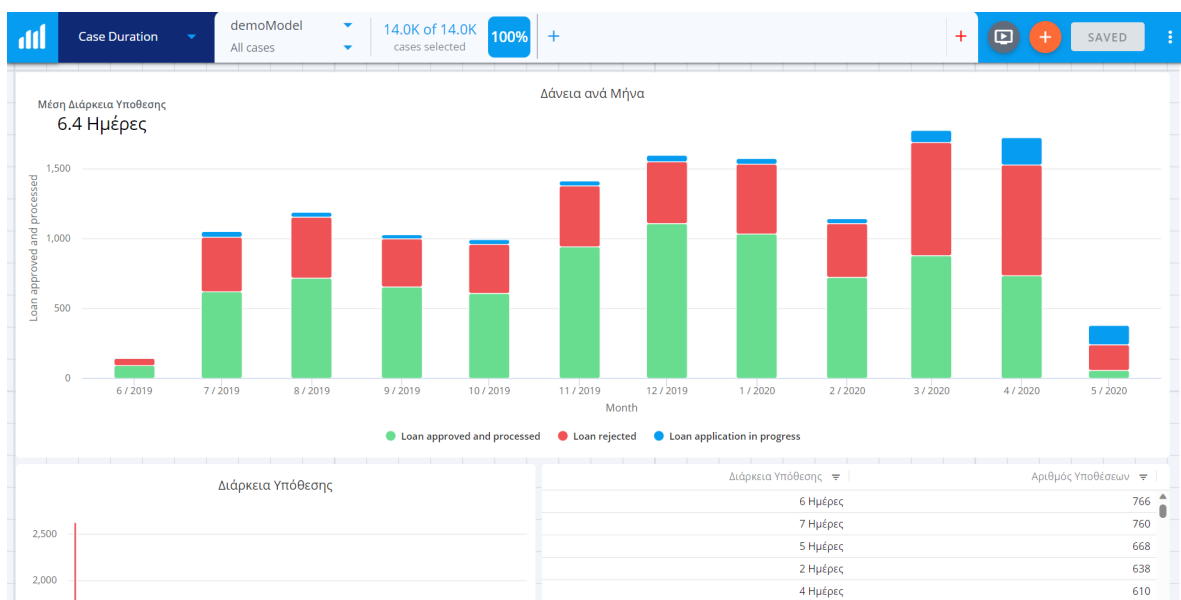
Εικόνα 4.16: Αναφορά για τις Υποθέσεις που αφορούν Δάνεια που έχουν απορριφθεί

Η αναφορά αυτή περιέχει ένα Γράφημα για την κατάσταση Δανείου ανά μήνα και μία ένδειξη για την μέση διάρκεια Υπόθεσης (Εικόνα 4.18). Περνώντας τον κέρσορα πάνω από κάθε μπάρα του γραφήματος μπορεί κανείς να δει τον ακριβή αριθμό Υποθέσεων ανά κατηγορία. Επίσης η αναφορά περιέχει ένα γράφημα για την διάρκεια των Υποθέσεων για τα εγκεκριμένα και απορριφθέντα δάνεια και έναν πίνακα για τον αριθμό Υποθέσεων ανά Διάρκεια (Εικόνα 4.19).

Στο Διάγραμμα για την Διάρκεια Υπόθεσης βλέπουμε ότι οι Υποθέσεις με εγκεκριμένα δάνεια υπερβαίνουν σε αριθμό τις Υποθέσεις με απορριφθέντα δάνεια όσο η διάρκεια μιας Υπόθεσης αυξάνεται. Επιλέγοντας με τον κέρσορα κάποιο σημείο του γραφήματος μπορούμε

Υποκ	Ποσοστό	Αριθμός Υποθέσεων
New Loan Application → Loan Rejected	35.5 %	1,815
New Loan Application → Process Loan Application → B2C First Approval → Loa...	27.3 %	1,397
New Loan Application → Process Loan Application → Loan Rejected	9.2 %	473
New Loan Application → Process Loan Application → B2B First Approval → Loa...	7.6 %	387
New Loan Application → Process Loan Application → Application Changed → P...	4.1 %	211
New Loan Application → Process Loan Application → Application Changed → P...	1.3 %	68
New Loan Application → Loan Rejected → Customer Asks for Further Assesme...	1.2 %	63
New Loan Application → Process Loan Application → Application Changed → P...	1.2 %	60
New Loan Application → Process Loan Application → B2C First Approval → App...	1.0 %	53
New Loan Application → Process Loan Application → B2C First Approval → Loa...	0.7 %	35
New Loan Application → Process Loan Application → B2C First Approval → App...	0.6 %	29
New Loan Application → Process Loan Application → Process Loan Application ...	0.5 %	28
New Loan Application → Process Loan Application → Application Changed → L...	0.5 %	26
New Loan Application → Process Loan Application → Application Changed → P...	0.4 %	23
New Loan Application → Process Loan Application → External Consultant → Lo...	0.4 %	23
New Loan Application → Process Loan Application → Process Loan Application ...	0.4 %	19
New Loan Application → Process Loan Application → B2B First Approval → App...	0.3 %	16
New Loan Application → Process Loan Application → B2B First Approval → Loa...	0.3 %	15
New Loan Application → Loan Rejected → Customer Asks for Further Assesme...	0.3 %	15
New Loan Application → Process Loan Application → Application Changed → C...	0.3 %	14
New Loan Application → Process Loan Application → B2C First Approval → App...	0.3 %	14
New Loan Application → Process Loan Application → Process Loan Application ...	0.2 %	10
New Loan Application → Process Loan Application → B2C First Approval → Loa...	0.2 %	10
New Loan Application → Process Loan Application → Loan Rejected → Custom...	0.2 %	9
New Loan Application → Process Loan Application → Application Changed → P...	0.2 %	8
New Loan Application → Process Loan Application → B2C First Approval → Loa...	0.2 %	8
New Loan Application → Loan Rejected → Customer Asks for Further Assesme...	0.2 %	8
New Loan Application → Process Loan Application → B2C First Approval → App...	0.2 %	8

Εικόνα 4.17: Πίνακας της Αναφοράς για τα Δάνεια που έχουν απορριφθεί

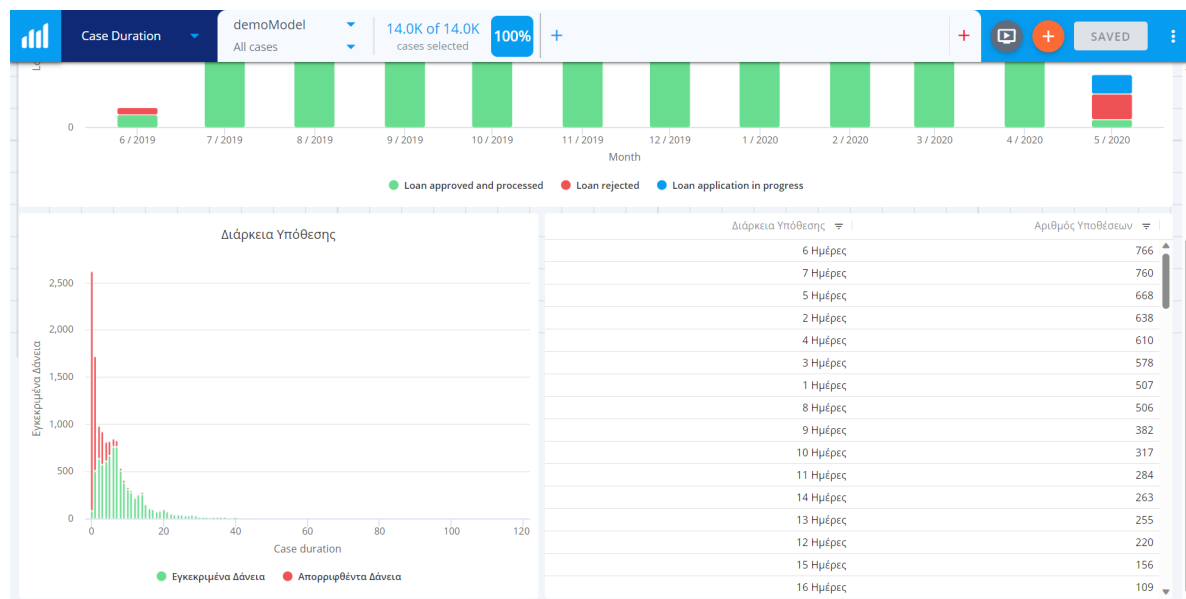


Εικόνα 4.18: Αναφορά για την Διάρκεια των Υποθέσεων (μέρος 1)

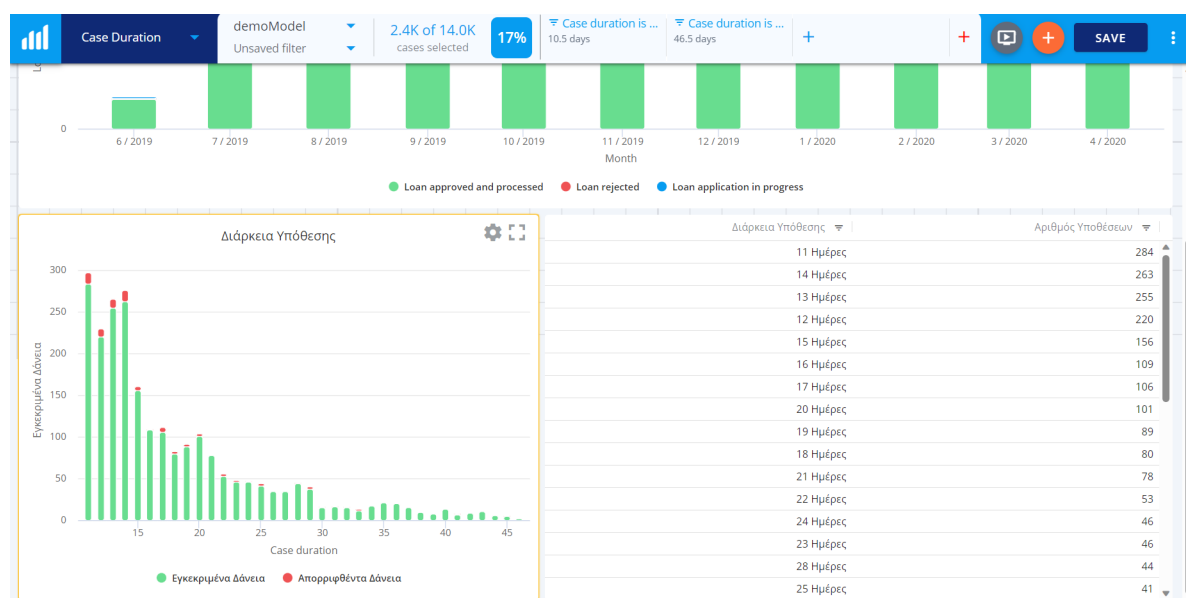
να εξάγουμε λεπτομερέστερα πληροφορίες για το σημείο αυτό.

Στην Εικόνα 4.20 έχουν επιλεγεί οι Υποθέσεις με Διάρκεια 11 ως 46 Ημέρες. Σε αυτές τις Υποθέσεις βλέπουμε ότι η κατάληξη των δανείων είναι στις περισσότερες περιπτώσεις θετική.

Η διάρκεια μίας Υπόθεσης είναι ένα χαρακτηριστικό που δεν υπάρχει στο Αρχείο Καταγραφής Υποθέσεων, μπορεί όμως να υπολογιστεί μέσω του Αρχείου Καταγραφής Γεγονότων αφαιρώντας την χρονική στιγμή έναρξης του Γεγονότος Έναρξης μιας Υπόθεσης από την χρονική στιγμή έναρξης του Γεγονότος Λήξης της ίδιας Υπόθεσης.



Εικόνα 4.19: Αναφορά για την Διάρκεια των Υποθέσεων (μέρος 2)



Εικόνα 4.20: Αναφορά για την Διάρκεια των Υποθέσεων (με φίλτρο διάρκειας 11 ως 46 ημέρες)

4.3.3 Ανάλυση Βασικών Αιτιών για Υποθέσεις με Καθυστέρηση

Όπως φαίνεται στην αναφορά της Εικόνας 4.18 η μέση Διάρκεια Υποθέσεων είναι 6.4 Ημέρες. Το εργαλείο QPR Process Analyzer παρέχει την δυνατότητα αναλύσεων με την ονομασία «Ανάλυση Βασικών Αιτιών» "Root Causes Analysis". Ωστόσο στην πραγματικότητα πρόκειται για την εύρεση παραγόντων που σχετίζονται με το εκάστοτε αποτέλεσμα, χωρίς να αποτελούν κατ' ανάγκη αιτίες του. Όπως αναφέρεται και στην Ενότητα 2.3 συχνά γίνεται σύγχυση της συσχέτισης με την αιτιότητα, κάτι που συμβαίνει και στην μέθοδο που χρησιμοποιεί το εργαλείο QPR Process Analyzer

Παραμένει όμως ενδιαφέρουσα η ανάλυση των κυριότερων παραγόντων που σχετίζονται με την καθυστέρηση μίας Υπόθεσης.

Χαρακτηριστικό Υπόθεσης	Τιμή Χαρακτηριστικού	Αριθμός Υποθέσεων	Μέση διάρκεια Υποθέσεων	Επίπεδο Συσχέτισης
Loan Status	Loan approved and processed	8,196	9.4	23.6 %
Loan rejection reason		8,914	9.1	23.0 %
Business Unit	Merchant	6,880	9.1	17.9 %
SLA	2.Out of Target	8,788	7.5	9.4 %
Product	Immediate loan (XX%)	2,384	8.8	5.6 %
Product	Car loan new (8%)	734	13.4	4.9 %
Product	Boat/ship loan (15%)	1,226	7.4	1.2 %
Location	Eugene	181	12.0	1.0 %
Location	Hong Kong	1,577	6.9	0.9 %
Product	Other consumer (high rate)	66	16.3	0.6 %
Location	Garland	69	15.1	0.6 %
Location	San Angelo	101	11.9	0.5 %
Location	Atlanta	115	11.0	0.5 %
Location	Shrewport	291	8.1	0.5 %
Location	Milwaukee	95	11.5	0.5 %
Location	Fullerton	65	13.0	0.4 %
Location	Fort Lauderdale	185	8.5	0.4 %
Location	Riverside	120	9.7	0.4 %
Location	Austin	117	9.8	0.4 %
Location	Dallas	513	7.1	0.4 %

Εικόνα 4.21: Αναφορά για τους παράγοντες που σχετίζονται με την καθυστέρηση μίας Υπόθεσης

Στην αναφορά που φαίνεται στην Εικόνα 4.21 παρουσιάζεται ένας πίνακας με τους παράγοντες που σχετίζονται με την καθυστέρηση μίας Υπόθεσης. Στην πρώτη στήλη φαίνεται το όνομα του Χαρακτηριστικού Υπόθεσης και στην δεύτερη η τιμή του, η οποία σχετίζεται με την καθυστέρηση. Στην τρίτη στήλη αναγράφεται ο αριθμός Υποθέσεων του Αρχείου Καταγραφής με την τιμή του χαρακτηριστικού που υποδηλώνουν οι 2 πρώτες στήλες. Στην τέταρτη στήλη φαίνεται η μέση διάρκεια Υπόθεσης (σε ημέρες) για τις Υποθέσεις με τα χαρακτηριστικά αυτά και τέλος στην πέμπτη στήλη φαίνεται το επίπεδο συσχέτισης μεταξύ της εκάστοτε τιμής χαρακτηριστικού και της καθυστέρησης μια Υπόθεσης.

Φαίνεται ότι το υψηλότερο επίπεδο συσχέτισης (23.6%) είναι μεταξύ της Έγκρισης ενός Δανείου και της καθυστέρησης της αντίστοιχης Υπόθεσης, με μέση διάρκεια των Υποθέσεων που αφορούν εγκεκριμένα δάνεια 9.4 Ημέρες.

Ακολουθούν με επίπεδο συσχέτισης 23% οι υποθέσεις στις οποίες δεν αναγράφεται λόγος απόρριψης δανείου, οι οποίες στην πραγματικότητα αποτελούν τις υποθέσεις στις οποίες είτε το δάνειο εγκρίθηκε είτε οι Υποθέσεις βρίσκονται σε εξέλιξη. Όπως είδαμε σε προηγούμενες αναφορές, η διάρκεια Υποθέσεων με δάνεια που απορρίφθηκαν είναι συνήθως μικρότερη από ότι των Υποθέσεων με δάνεια που εγκρίθηκαν. Συνεπώς είναι λογικό η μέση διάρκεια Υποθέσεων με δάνεια που εγκρίθηκαν να είναι μεγαλύτερη από την μέση διάρκεια όλων των υποθέσεων. Αυτό όμως δεν σημαίνει ότι υπάρχει σχέση αιτιότητας μεταξύ της έγκρισης ενός δανείου και της καθυστέρησης της Υπόθεσης.

Ενδιαφέρον παρουσιάζει η τρίτη γραμμή του πίνακα που αφορά την τιμή Merchant του χαρακτηριστικού Business Unit, δηλαδή την εμπορική επιχειρηματική μονάδα της τράπεζας. Οι Υποθέσεις που διεξήχθησαν σε αυτήν την μονάδα φαίνεται να σχετίζονται σε επίπεδο 17.9% με την καθυστέρηση των Υποθέσεων, ενώ η μέση διάρκεια Υπόθεσης στην περίπτωση αυτή είναι 9.1 Ημέρες. Το αποτέλεσμα αυτό δεν πρόκυπτε από προηγούμενες αναλύσεις και η περαιτέρω ανάλυσή του μπορεί να δώσει χρήσιμες πληροφορίες στην τράπεζα. Και σε αυτή την περίπτωση ωστόσο η ύπαρξη συσχέτισης δεν συνεπάγεται αιτιότητα.

Κεφάλαιο 5

Συμπεράσματα - Μελλοντική Έρευνα

Στα προηγούμενα Κεφάλαια της εργασίας έγινε περιγραφή των βασικών εννοιών και μεθόδων που σχετίζονται με την Εξόρυξη Διεργασιών και μία σύντομη παρουσίαση ορισμένων αναλύσεων ως αποτέλεσμα εφαρμογής των τεχνικών εξόρυξης διεργασιών. Με αφετηρία ένα Αρχείο Καταγραφής Γεγονότων εμπλουτισμένο με ένα Αρχείο Καταγραφής Υποθέσεων παρουσιάστηκε το μοντέλο της διεργασίας που αυτά περιγράφουν, ορισμένες αναλύσεις συχνοτήτων και οι παράγοντες που σχετίζονται περισσότερο με την καθυστέρηση Υποθέσεων. Τα δύο τελευταία αποτελέσματα θα μπορούσαν να βρεθούν και με παραδοσιακές μεθόδους στατιστικής, αλλά αυτό θα απαιτούσε περισσότερο χρόνο και ο τρόπος δεν θα ήταν τόσο άμεσος.

Τα πλεονεκτήματα της εξόρυξης διεργασιών ωστόσο δεν περιορίζονται σε αυτά. Σε μελλοντική έρευνα μπορεί να μελετηθεί σε μεγαλύτερο βάθος η διεργασία που μας απασχόλησε, χρησιμοποιώντας ένα πιο εμπλουτισμένο αρχείο ή προσθέτοντας ορισμένα χαρακτηριστικά μέσω υπολογισμών (όπως έγινε με την διάρκεια υπόθεσης). Επιπλέον μπορούν να ανακαλυφθούν και να μελετηθούν σε μεγαλύτερο βάθος οι δυσλειτουργίες της Διεργασίας και να εξαχθούν συμπεράσματα για τις αποκλίνουσες συμπεριφορές, δηλαδή τα Ίχνη του Αρχείου Καταγραφής που αποκλίνουν από την συνήθη συμπεριφορά. Τέλος, μπορούν να εφαρμοστούν οι μέθοδοι ανάλυσης βασικών αιτιών που παρουσιάζονται στην Ενότητα 2.3, ώστε να ανιχνευθούν οι παράγοντες που πραγματικά αποτελούν αιτίες για την καθυστέρηση των Υποθέσεων.

Παραρτήματα

Κώδικες του QPR Process Analyzer

Listing A.1: Πίνακας Γεγονότων Έναρξης

```
{
  "Dimensions": [
    {
      "Name": "dimension0",
      "Expression": "FirstEvent.TypeName"
    }
  ],
  "Values": [
    {
      "Name": "measure0",
      "Expression": "Count(_)"
    }
  ],
  "Ordering": [
    {
      "Name": "measure0",
      "Direction": "Descending"
    }
  ],
  "AggregateOthers": false,
  "Root": "Cases",
  "ModelId": 40650,
  "EnableResultCaching": true,
  "QueryIdentifier": "T100_T101_12620",
  "ContextType": "eventlog",
  "CancelEarlierQueriesWithIdentifier": true
}
```

Listing A.2: Πίνακας Γεγονότων Λήξης

```
{
  "Dimensions": [
```

```

    {
        "Name": "dimension0",
        "Expression": "LastEvent.TypeName"
    }
],
"Values": [
    {
        "Name": "measure0",
        "Expression": "Count(_)"
    }
],
"Ordering": [
    {
        "Name": "measure0",
        "Direction": "Descending"
    }
],
"AggregateOthers": false,
"Root": "Cases",
"ModelId": 40650,
"EnableResultCaching": true,
"QueryIdentifier": "T102_T103_12620",
"ContextType": "eventlog",
"CancelEarlierQueriesWithIdentifier": true
}

```

Listing A.3: Πίνακας Ίχνη Υποθέσεων

```

{
  "Dimensions": [
    {
      "Name": "dimension0",
      "Expression": "StringJoin(\" → \",
        Variation.EventTypes.Name)"
    }
  ],
  "Values": [
    {
      "Name": "measure0",
      "Expression": "Count(_)"
    },
    {
      "Name": "measure1",
      "Expression": "Count(_)/allCasesCount*100"
    }
  ]
}

```



```

    }
  ],
  "Ordering": [
    {
      "Name": "measure0",
      "Direction": "Descending"
    },
    {
      "Name": "measure1",
      "Direction": "Descending"
    }
  ],
  "AggregateOthers": false,
  "Root": "let sampleSize = null;let allCasesCount =
CaseCount. If (sampleSize!=null&&>sampleSize , sampleSize , _); Cases",
  "ModelId": 40650,
  "EnableResultCaching": true,
  "QueryIdentifier": "T104_T105_12620",
  "ContextType": "eventlog",
  "CancelEarlierQueriesWithIdentifier": true
}

```

Listing A.4: "Γράφημα Ανά Κατάσταση Αίτησης"

```

{
  "Dimensions": [
    {
      "Name": "dimension0",
      "Expression": "Attribute(\"Loan Status\")"
    }
  ],
  "Values": [
    {
      "Name": "measure0",
      "Expression": "Count(_)/allCasesCount*100"
    }
  ],
  "Ordering": [],
  "AggregateOthers": false,
  "Root": "let sampleSize = null;let allCasesCount =
CaseCount. If (sampleSize!=null&&>sampleSize , sampleSize , _);
Cases",
  "ModelId": 40650,
  "EnableResultCaching": true,

```

```

"QueryIdentifier": "T36_T37_T35",
"ContextType": "eventlog",
"CancelEarlierQueriesWithIdentifier": true
}

```

Listing A.5: Πίνακας Κατάσταση Αίτησης

```

{
  "Dimensions": [
    {
      "Name": "dimension0",
      "Expression": "Attribute(\"Month\")",
      "DatetimeTruncation": "month"
    }
  ],
  "Values": [
    {
      "Name": "measure0",
      "Expression": "Count(_)"
    },
    {
      "Name": "measure1",
      "Expression": "Count(_.Where(Attribute(\"Loan Status\")==\"Loan approved and processed\"))"
    },
    {
      "Name": "measure2",
      "Expression": "Count(_.Where(Attribute(\"Loan Status\")==\"Loan application in progress\"))"
    },
    {
      "Name": "measure3",
      "Expression": "Count(_.Where(Attribute(\"Loan Status\")==\"Loan rejected\"))"
    },
    {
      "Name": "measure4",
      "Expression":
"Count(_.Where(Attribute(\"SLA\")==\"1.InTarget\"))"
    },
    {
      "Name": "measure5",
      "Expression":
"Count(_.Where(Attribute(\"SLA\")==\"2.Out of Target\"))"
    }
  ]
}

```

```

    }
  ],
  "Ordering": [
    {
      "Name": "dimension0",
      "Direction": "Ascending"
    }
  ],
  "AggregateOthers": false,
  "Root": "Cases",
  "ModelId": 40650,
  "EnableResultCaching": true,
  "QueryIdentifier": "T58_T59_T35",
  "ContextType": "eventlog",
  "CancelEarlierQueriesWithIdentifier": true
}

{
  "Dimensions": [
    {
      "Name": "dimension0",
      "Expression": "Attribute(\"Loan rejection reason\")"
    }
  ],
  "Values": [
    {
      "Name": "measure0",
      "Expression": "Count(_)/allCasesCount*100"
    }
  ],
  "Ordering": [
    {
      "Name": "dimension0",
      "Direction": "Ascending"
    }
  ],
  "AggregateOthers": false,
  "Root": "let sampleSize = null;let allCasesCount =
CaseCount. If (sampleSize!=null&&_>sampleSize, sampleSize, _);
Cases",
  "ModelId": 40650,
  "Filter": {
    "Items": [

```

```

        {
            "Type": "IncludeCases",
            "Items": [
                {
                    "Type": "CaseAttributeValue",
                    "Attribute": "Loan Status",
                    "StringifiedValues": [
                        "0Loan rejected"
                    ]
                }
            ]
        }
    ],
    "EnableResultCaching": true,
    "QueryIdentifier": "T87_T88_12631",
    "ContextType": "eventlog",
    "CancelEarlierQueriesWithIdentifier": true
}

```

Listing A.6: "Πίνακας απορριφθέντων Δανείων ανά Τχνος

```

{
    "Dimensions": [
        {
            "Name": "dimension0",
            "Expression": "StringJoin(\"      \", Variation.EventTypes.Name)"
        }
    ],
    "Values": [
        {
            "Name": "measure0",
            "Expression": "Count(_)/allCasesCount*100"
        },
        {
            "Name": "measure1",
            "Expression": "Count(_)"
        }
    ],
    "Ordering": [
        {
            "Name": "measure0",
            "Direction": "Descending"
        },
    ]
}

```

```

        {
            "Name": "measure1",
            "Direction": "Descending"
        }
    ],
    "AggregateOthers": false,
    "Root": "let sampleSize = null;let allCasesCount =
CaseCount.If(sampleSize!=null&&_>sampleSize,sampleSize,_);Cases",
    "ModelId": 40650,
    "Filter": {
        "Items": [
            {
                "Type": "IncludeCases",
                "Items": [
                    {
                        "Type": "CaseAttributeValue",
                        "Attribute": "Loan Status",
                        "StringifiedValues": [
                            "0Loan rejected"
                        ]
                    }
                ]
            }
        ]
    },
    "EnableResultCaching": true,
    "QueryIdentifier": "T99_T100_12631",
    "ContextType": "eventlog",
    "CancelEarlierQueriesWithIdentifier": true
}

{
    "Dimensions": [
        {
            "Name": "dimension0",
            "Expression": "Attribute(\"Month\")",
            "DatetimeTruncation": "month"
        }
    ],
    "Values": [
        {
            "Name": "measure0",
            "Expression": "Count(_.Where(Attribute(\"Loan Status\")==

```

```

        \ "Loan approved and processed \")"
    },
    {
        "Name": "measure1",
        "Expression": "Count(_.Where(Attribute(\ "Loan Status \")==
            \ "Loan rejected \"))"
    },
    {
        "Name": "measure2",
        "Expression": "Count(_.Where(Attribute(\ "Loan Status \")==
            \ "Loan application in progress \"))"
    }
],
"Ordering": [],
"AggregateOthers": false,
"Root": "Cases",
"ModelId": 40650,
"EnableResultCaching": true,
"QueryIdentifier": "T111_T112_12626",
"ContextType": "eventlog",
"CancelEarlierQueriesWithIdentifier": true
}

```

Listing A.7: "Γράφημα Διάρκεια Υπόθεσης"

```

{
    "Dimensions": [
        {
            "Name": "dimension0",
            "Expression": "Duration?.TotalDays",
            "NumberPrecision": 0
        }
    ],
    "Values": [
        {
            "Name": "measure0",
            "Expression": "Count(_.Where(Attribute(\ "Loan Status \")==
                \ "Loan approved and processed \"))"
        },
        {
            "Name": "measure1",
            "Expression": "Count(_.Where(Attribute(\ "Loan Status \")==
                \ "Loan rejected \"))"
        }
    ]
}

```

```

],
"Ordering": [],
"AggregateOthers": false,
"Root": "Cases",
"ModelId": 40650,
"EnableResultCaching": true,
"QueryIdentifier": "T113_T114_12626",
"ContextType": "eventlog",
"CancelEarlierQueriesWithIdentifier": true
}

```

Listing A.8: "Πίνακας Διάρκεια Υπόθεσης"

```

{
  "Dimensions": [
    {
      "Name": "dimension0",
      "Expression": "Duration?.TotalDays",
      "NumberPrecision": 0
    }
  ],
  "Values": [
    {
      "Name": "measure0",
      "Expression": "Count(_.Where(Attribute(\"Loan Status\")==
        \"Loan approved and processed\"))"
    }
  ],
  "Ordering": [
    {
      "Name": "measure0",
      "Direction": "Descending"
    }
  ],
  "AggregateOthers": false,
  "Root": "Cases",
  "ModelId": 40650,
  "EnableResultCaching": true,
  "QueryIdentifier": "T115_T116_12626",
  "ContextType": "eventlog",
  "CancelEarlierQueriesWithIdentifier": true
}

```

Listing A.9: "Πίνακας με τους παράγοντες που σχετίζονται με την καθυστέρηση μίας Υπόθεσης

```

{
  "Dimensions": null,
  "Values": [
    {
      "Name": "dimension0",
      "Expression": "Column(\"AttributeName\")"
    },
    {
      "Name": "dimension1",
      "Expression": "Column(\"AttributeValue\")"
    },
    {
      "Name": "dimension2",
      "Expression": "Column(\"GroupCaseCount\")"
    },
    {
      "Name": "dimension3",
      "Expression": "Column(\"GroupMeasureValue\")"
    },
    {
      "Name": "dimension4",
      "Expression": "Column(\"ContributionPercentage\")"
    }
  ],
  "Ordering": [],
  "AggregateOthers": false,
  "Root":
  "let el = _;\nlet cAttributes =
  el.CaseAttributes.Where(Datatype.In([\\"String\\",\\"Integer\\",\\"Boolean\\"]))
  &&
  UniqueCount < 1000
  &&
  (Count([\\"Business Unit\\",\\"Cost\\",\\"Loan Status\\",\\"Loan rejection \n
  reason\\",\\"Location\\",\\"Product\\",\\"SLA\\",\\"User\\"])=0 \n
  || \n
  Name.In([\\"Business Unit\\",\\"Cost\\",\\"Loan Status\\",\\"Loan rejection \n
  reason\\",\\"Location\\",\\"Product\\",\\"SLA\\",\\"User\\"]));\n
  let measureTotal = Average(el.Cases.(Duration?.TotalDays));\n
  let result = ToDataFrame([], [\"AttributeValue\", \"GroupCaseCount\",\n
  \"GroupMeasureValue\", \"AttributeName\"]);

```

```

for (let i = 0; i < Count(cAttributes); i++) \n
{\n  let dataPart = ToDataFrame(e1.Cases.[ Attribute(cAttributes[i]), \n
(Duration?.TotalDays)], [\"AttributeValue\", \"CaseMeasure\"])\n
.GroupBy([\"AttributeValue\"])\n
.Aggregate(\n      [\"GroupCaseCount\": \"AttributeValue\", \n
\"GroupMeasureValue\": \n
\"CaseMeasure\"], \n
[\"Count\", \"Average\"]\n
).SetColumns([\n
\"AttributeName\": () => cAttributes[i].Name\n
]);\n
result = result.append(dataPart);\n
}\nresult = result.SetColumns([\n
\"DifferenceToOverall\": () => GroupMeasureValue – measureTotal, \n
\"Contribution\": () => DifferenceToOverall * GroupCaseCount, \n
\"ContributionAbs\": () => Abs(DifferenceToOverall * GroupCaseCount)\n]);\n
let contributionSum = Sum(result.ContributionAbs) / \n
2;\nresult\n.OrderByColumns([\"ContributionAbs\"], [false])\n
.Head(null ?? 2147483647)\n
.OrderByColumns([\"Contribution\"], [false])\n
.SetColumns([\n  \"TotalMeasureValue\": () => measureTotal, \n
\"ContributionPercentage\": () => If(contributionSum == 0, 0, Contribution \n
/ contributionSum * 100) ]);",
      "ModelId": 40650,
      "EnableResultCaching": true,
      "QueryIdentifier": "T119_T120_12632",
      "ContextType": "eventlog",
      "CancelEarlierQueriesWithIdentifier": true
}

```


Βιβλιογραφία

- [1] Christian Günther και Wil Aalst. *Fuzzy Mining - Adaptive Process Simplification Based on Multi-perspective Metrics*. τόμος 4714, σελίδες 328–343, 2007.
- [2] Wil Aalst, Arya Adriansyah, Ana Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose R.P., Peter Brand, Ronald Brandtjen, Joos Buijs, Andrea Burattin, Josep Carmona, Malú Castellanos, Jan Claes, Jonathan Cook, Nicola Costantini, Francisco Curbera, Ernesto Damiani, Massimiliano Leoni και Moe Wynn. *Process Mining Manifesto*. τόμος 99, σελίδες 169–194, 2011.
- [3] W. M. P. Vande Aalst και A. J. M. M. Weijters. *Process mining: A research agenda*. *Computers in Industry*, 53(3):231 – 244, 2004.
- [4] Wil Aalst. *Process Mining: A 360 Degree Overview*, σελίδες 3–34. 2022.
- [5] *AIMultiple: In-depth, transparent insights for data-driven decision making*. <https://aimultiple.com/about>. Αρχειοθετημένο: 2023-07-16.
- [6] *Top 44 Process Mining Use Cases Applications in 2023*. <https://research.aimultiple.com/process-mining-use-cases/#healthcare>. Αρχειοθετημένο: 2023-07-16.
- [7] W. M. P. Vande Aalst. *Process Mining: Overview and Opportunities*. *ACM Transactions on Management Information Systems*, 3(2):1 – 17, 2012.
- [8] Wil Aalst. *Foundations of Process Discovery*, σελίδες 37–75. 2022.
- [9] Kenneth H. Rosen. *Διακριτά Μαθηματικά και Εφαρμογές*. Τζιόλα, Αθήνα, 8η έκδοση, 2019.
- [10] Frank Fox, Vishal Aggarwal, Helen Whelton και Owen Johnson. *A Data Quality Framework for Process Mining of Electronic Health Record Data*. 2018.
- [11] Marcus Dees, Massimiliano Leoni και Felix Mannhardt. *Enhancing Process Models to Improve Business Performance: A Methodology and Case Studies*. 2017.
- [12] Philip Weber, Behzad Bordbar και Peter Tino. *A Framework for the Analysis of Process Mining Algorithms*. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, 43:303–317, 2013.
- [13] Asst Esmita και Esmita Gupta. *Process Mining A Comparative Study*. 2014.
- [14] A. Weijters, Wil Aalst και Alves Medeiros. *Process Mining with the Heuristics Miner-algorithm*, τόμος 166. 2006.

- [15] Asst Esmita και Esmita Gupta. *PROCESS MINING ALGORITHMS*. 2014.
- [16] Wil Aalst, Ana Medeiros και A. Weijters. *Genetic Process Mining*. τόμος 14, σελίδες 48–69, 2005.
- [17] A. Medeiros, A. Weijters και Wil Aalst. *Genetic Process Mining: A Basic Approach and Its Challenges*. τόμος 3812, σελίδες 203–215, 1970.
- [18] Sander Leemans, Dirk Fahland και Wil Aalst. *Scalable process discovery and conformance checking*. *Software Systems Modeling*, 17, 2018.
- [19] M.L. Eck, van, X. Lu, S.J.J. Leemans και W.M.P. Aalst, van der. *PM2 : a Process Mining Project Methodology*. *Advanced Information Systems Engineering : 27th International Conference, CAiSE 2015, Stockholm, Sweden, June 8-12, 2015. Proceedings*J. Zdravkovic, M. Kirikova και P. Johannesson, επιμελητές, Lecture Notes in Computer Science, σελίδες 297–313, Germany, 2015. Springer. 27η Ιντερνατιοναλ ὄνφερενσε ὀν Αδανσεδ Ινφορματιον Σψοτερμς Ενγινεερινγ (ΑιΣΕ 2015), ΑιΣΕ 2015 · ὄνφερενσε δατε: 08-06-2015 Τηρουγη 12-06-2015.
- [20] Josep Carmona, Boudewijn Dongen και Matthias Weidlich. *Conformance Checking: Foundations, Milestones and Challenges*, σελίδες 155–190. 2022.
- [21] Arya Adriansyah. *Aligning observed and modeled behavior*. 2014.
- [22] Michael Becker και Ralf Laue. *A comparative survey of business process similarity measures*. *Computers in Industry*, 63(2):148–167, 2012. Managing Large Collections of Business Process Models.
- [23] Remco Dijkman, Boudewijn Dongen, Marlon Dumas, Luciano García-Bañuelos, Matthias Kunze, Henrik Leopold, Jan Mendling, Reina Uba, Matthias Weidlich, Mathias Weske και Zhiqiang Yan. *A Short Survey on Process Model Similarity*, σελίδες 421–427. 2013.
- [24] Bart Hompes, Abderrahmane Maaradji, Marcello La Rosa, Marlon Dumas, Joos Buijs και Wil Aalst. *Discovering Causal Factors Explaining Business Process Performance Variation*. σελίδες 177–192, 2017.
- [25] Greg Van Houdt, Benoît Depaire, Niels Martin και Carlos Mosquera. *Root Cause Analysis in Process Mining with Probabilistic Temporal Logic*. 2021.
- [26] Suriadi Suriadi, Wil Aalst και Arthur Ter. *Root Cause Analysis with Enriched Process Logs*. τόμος 132, σελίδες 174–186, 2013.
- [27] Bart Hompes, Abderrahmane Maaradji, Marcello La Rosa, Marlon Dumas, Joos Buijs και Wil Aalst. *Discovering Causal Factors Explaining Business Process Performance Variation*. σελίδες 177–192, 2017.
- [28] C.W.J. Granger. *Some recent development in a concept of causality*. *Journal of Econometrics*, 39(1):199–211, 1988.

- [29] Mahnaz Qafari και Wil Aalst. *Case Level Counterfactual Reasoning in Process Mining*, 2021.
- [30] *ProM Tools - The Process Mining Framework*, 2023. [Online; accessed 7 -September-2023].
- [31] B. F.van Dongen, A. K. A.de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters και W. M. P.van der Aalst. *The ProM Framework: A New Era in Process Mining Tool Support. Applications and Theory of Petri Nets 2005* Gianfranco Ciardo και Philippe Darondeau, επιμελητές, σελίδες 444-454, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.
- [32] *ProM Documentation - ProM Tools*, 2023. [Online; accessed 7 -September-2023].
- [33] QPR ProcessAnalyzer Wiki. *Introduction to QPR ProcessAnalyzer – QPR ProcessAnalyzer Wiki*, 2021. [Online; accessed 12-July-2023].
- [34] QPR ProcessAnalyzer Wiki. *QPR ProcessAnalyzer System Architecture – QPR ProcessAnalyzer Wiki*, 2023. [Online; accessed 12-July-2023].

Συντομογραφίες - Αρκτικόλεξα - Ακρωνύμια

βλπ

κ.λπ.

κ.ο.κ

βλέπε

και λοιπά

και ούτω καθεξής

Απόδοση ξενόγλωσσων όρων

Απόδοση

- Αιτιώδεις Παράγοντες
- Ανακάλυψη
- Ανάλυση Βασικών Αιτιών
- Ανάλυση Διεργασιών
- Αντιφατικά Στιγμιότυπα
- Αξιοποιήσιμα Γνωρίσματα Κατάστασης
- Αποδεκτό Δίκτυο Petri
- Αρχείο Καταγραφής Γεγονότων
- Γεγονός
- Γενετικός Τελεστής
- Γράφημα Άμεσης Ακολουθίας
- Γράφημα Ανάλυσης
- Διασταύρωση
- Δίκτυο Petri
- Δραστηριότητα
- Ελάχιστη Ασαφής Εμπιστοσύνη
- Ελάχιστη Ασαφής Υποστήριξη
- Έλεγχος Συμμόρφωσης
- Ελιτισμός
- Εξόρυξη Δεδομένων
- Εξόρυξη Διεργασιών
- Επισημασμένο Δίκτυο Petri
- Ευθυγράμμιση
- Ίχνος
- Μεταλλαγή
- Μετατοπίσεις Ευθυγράμμισης
- Μοντέλο διαρθρωτικών εξισώσεων
- Μοντελοποίηση Διεργασιών
- Οντολογία του Τομέα
- Σήμανση Έναρξης
- Σήμανση Λήξης
- Στιγμιότυπο της Διεργασίας
- Σύμβολα
- Συνάρτησης Απόδοσης

Ξενόγλωσσος όρος

- Causal Factors
- Discover
- Root Causes Analysis
- Process Analysis
- Counterfactual Instances
- Actionable Situation Features
- Accepting Petri Net
- Event Log
- Even
- Genetic Operator
- Directly-follows graph
- Decomposition Graph
- Crossover
- Petri Net
- Activity
- Minimum Fuzzy Confidence
- Minimum Fuzzy Support
- Conformance Checking
- Elitism
- Data Mining
- Process Mining
- Labeled Petri Net
- Alignment
- Trace
- Mutation
- Alignment Moves
- Structural Equation Model
- Process Modeling
- Domain Ontology
- Initial Marking
- Final Marking
- Process Instance
- Tokens
- Performance function

- | | |
|------------------------------------|-----------------------------------|
| • Σύνολο Αντιφατικών Αιτιολογήσεων | Set of Counterfactual Explanation |
| • Υπόθεση | Case |
| • Υπολογιστική Νοημοσύνη | computational intelligence |
| • Χαρακτηριστικό | Attribute |
| • Χρονική Σήμανση | Timestamp |