



ΕΘΝΙΚΟ ΜΕΤΣΟΒΕΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΚΑΤΑΣΚΕΥΗ ΥΠΟΓΕΙΩΝ ΕΡΓΩΝ»

**ΕΚΤΙΜΗΣΗ ΚΙΝΔΥΝΟΥ ΑΣΤΟΧΙΑΣ ΣΗΡΑΓΓΑΣ ΜΕ
ΕΦΑΡΜΟΓΗ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Εμμανουήλ Βενιέρης

Επιβλέπων: Ανδρέας Μπενάρδος

Αναπληρωτής Καθηγητής Ε.Μ.Π.

ΑΘΗΝΑ, ΙΟΥΝΙΟΣ 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΕΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΚΑΤΑΣΚΕΥΗ ΥΠΟΓΕΙΩΝ ΕΡΓΩΝ»

ΕΚΤΙΜΗΣΗ ΚΙΝΔΥΝΟΥ ΑΣΤΟΧΙΑΣ ΣΗΡΑΓΓΑΣ ΜΕ ΕΦΑΡΜΟΓΗ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Εμμανουήλ Βενιέρης

Επιβλέπων: Ανδρέας Μπενάρδος

Αναπληρωτής Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 21^η Ιουλίου 2021.

Α. Μπενάρδος
Αν. Καθηγητής Ε.Μ.Π

Β. Μαρίνος
Επ. Καθηγητής Ε.Μ.Π

Χ. Παρασκευοπούλου
Επ. Καθηγητής Leeds Univ. UK

ΑΘΗΝΑ, ΙΟΥΛΙΟΣ 2021

.....

Εμμανουήλ Φ. Βενιέρης

Διπλωματούχος Μηχανικός Εγκαταστάσεων Σχολής Ικάρων

Copyright © Εμμανουήλ Φ. Βενιέρης

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Αντικείμενο της παρούσας εργασίας είναι η ανάλυση δεδομένων περιστατικών αστοχίας σηράγγων, μέσω εφαρμογής μοντέλων μηχανικής μάθησης που ενσωματώνουν αλγόριθμους συσταδοποίησης. Η ανάλυση αποσκοπεί στην εξόρυξη προτύπων που δύναται να συσχετιστούν με συγκεκριμένες κατηγορίες κινδύνου αστοχίας, καθώς και στην αξιοποίηση των αποτελεσμάτων συσταδοποίησης, για την αξιολόγηση σεναρίων κατασκευής σηράγγων. Τα μοντέλα που αναπτύχθηκαν προσφέρουν τη δυνατότητα στο χρήστη να εισάγει τα χαρακτηριστικά μίας υπό εξέταση σήραγγας και να εξάγει αποτελέσματα που αφορούν στον κίνδυνο αστοχίας της και στην ομοιότητά της με σήραγγες που περιλαμβάνονται στη βάση δεδομένων. Επιπλέον, μπορούν να χρησιμοποιηθούν για την σύγκριση σεναρίων με πιθανοτικά κριτήρια. Τα αποτελέσματα της ανάλυσης ερμηνεύονται και τα μοντέλα αξιολογούνται ως προς την ακρίβεια, την αξιοπιστία και την αποτελεσματικότητά τους.

Abstract

The construction of underground infrastructure and in particular the construction of tunnels, in contrast to the construction of above the ground facilities, suffers from a large amount of uncertainty. To deal with this uncertainty a large expense must be put into the preliminary design in order to produce an adequate geotechnical model. Unfortunately, there is a limit to how much companies are willing to spend for the production of this model. One way to know how much it is worth spending, is by evaluating the risk of failure, given a set of parameters for an exact tunnelling project.

In this thesis, a pre-existing partially completed database of significant tunnel failures was processed, in order to retrieve a usable set of data (dataset TFD_21). The dataset contains 118 failure incidents and through the application of statistical techniques (such as PCA) it was ensured that it constitutes a representative sample of the database. The tunnels' attributes contained in the dataset are "length" (m), "diameter" (m), "overburden" (m), "tunnel use", "cutting technique" and "rock/soil". The first three are numerical attributes and the rest are categorical. Dataset TFD_21 was analyzed through the application of two machine learning models, for the purpose of revealing patterns that lead to certain risk of failure. The models can also be used in order to compare different tunnel construction scenarios based on the clustering results, using probabilistic criterions. The programming language used for the encoding of the models was Python and the application used to edit and run the code was Jupyter Notebook. In Figure 1, dataset TFD_21 is shown, when inserted for analysis in model TFM1.

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Thirre Tunnel	Road	NATM	5500	12.60	650.0	Rock
1	Lane cove Tunnel	Ventilation	NATM	130	8.00	17.0	Rock
2	Wienerwald tunnel, Vienna	Rail	NATM/D&B	13350	10.68	190.0	Rock
3	Lainzer railroad tunnel, Vienna	Rail	NATM	12800	5.55	26.0	Soil
4	Lambach Rail Tunnel	Rail	NATM	912	12.00	46.0	Rock
...
102	Seattle's Beacon Hill Light Rail	Rail	TBM	25000	6.40	7.0	Soil
103	Dulles airport, Washington	Walk	NATM	366	11.00	4.5	Soil
104	Hollywood section, LA Metro extension, Los Ang...	Metro	TBM	14000	6.70	25.0	Soil
105	Severomuisk railway tunnel, Baikal-Amur Line,	Rail	TBM	15343	4.56	1000.0	Rock
106	Haivan Pass Tunnel, Lang Co	Road	D&B	6280	11.90	890.0	Rock

107 rows x 7 columns

Figure 1. Dataset TFD_21 inserted for analysis in TFM1.

Models TFM1 and TFM2, in their first section (TFM1_A and TFM2_A), embody the clustering algorithms K-Means and K-Prototypes, respectively. The clusters are labeled according to each cluster's relative frequency, which is the proportion of one cluster's incidents by the total number of incidents. The application's clustering results are shown in Table 1 and the estimated patterns in Table 2. Figure 2 contains a graphical representation of the important attribute values.

<i>Cluster Label</i>	<i>Cluster's Relative Frequency</i>	
	<i>TFM1 (K-Means)</i>	<i>TFM1 (K-Prototypes)</i>
Low Frequency	15,0 %	4,7 %
Medium Frequency	19,6 %	22,4 %
High Frequency	21,5 %	27,1 %
Very High Frequency	43,9 %	45,8 %

Table 1. Cluster labels and clusters' relative frequencies.

<i>Feature</i>	<i>Low Frequency</i>	<i>Medium Frequency</i>	<i>High Frequency</i>	<i>Very High Frequency</i>
	<i>TFM1 (K-Means)</i>			
Cutting Technique	Conventional Methods	TBM	TBM	Conventional Methods
Length	< 7 km	< 7km	> 7km	> 7km
Diameter	> 10 m	< 10 m	< 10 m	> 10 m
Overburden	< 40 m	< 40 m	> 40 m	> 40 m
Rock/Soil	Soil	Soil	Rock	Rock
<i>TFM2 (K-Prototypes)</i>				
Cutting Technique	All methods	TBM	TBM	Conventional Methods
Length	< 2,5km	> 8km	< 8km	All lengths
Diameter	> 7m	< 10 m	< 10 m	> 10 m
Overburden	> 2000 m	> 40 m	< 40 m	All depths
Rock/Soil	Rock	Rock	Soil	Rock

Table 2. Estimated patterns related to each cluster (TFM1 and TFM2).



Figure 2. Graphical representation of the important attribute values for each cluster (TFM1 and TFM2)

After the interpretation, evaluation and combination of TFM1 and TFM2 results, the conclusion reached was that there are four (4) patterns related to the clusters, but the failure risk categories are three (3) due to the fact that an examined tunnel can fall to the medium

risk category by consorting with two of the four patterns. The patterns revealed and the corresponding risk categories are shown in Table 3.

<i>Feature</i>	<i>Patterns</i>			
	<i>Low Risk Cat.</i>	<i>Medium Risk Category</i>		<i>High Risk Cat.</i>
Cutting Technique	Conventional Methods	TBM	TBM	Conventional Methods
Length	< 7 km	< 7km	> 8km	> 7km
Diameter	> 10 m	< 10 m	< 10 m	> 10 m
Overburden	< 40 m	< 40 m	> 40 m	> 40 m
Rock/Soil	Soil	Soil	Rock	Rock

Table 3. Patterns revealed from combination or results and corresponding failure risk categories.

For the second section of the models (TFM1_B and TFM2_B), a methodology was developed in order to quantify the similarity of a given tunnel construction scenario to each cluster’s centroid and to the “nearest neighbor” incident, and based on these, estimate the risk category of the project and the relative failure probability. The term relative is used because the absolute failure probability cannot be calculated, due to the fact that there isn’t available data about all the tunnels that haven’t failed across the world. Therefore, this relative failure probability can only be used for the comparison of scenarios. The methodology developed, uses the distance and dissimilarity measures that K-Means and K-Prototypes algorithms also use, in order to calculate the percentage similarities. The input of the parameters in TFM1 model, for a test tunnel, is shown in Figure 3 and the results are shown in Figure 4.

```
# INPUT TEST TUNNEL FEATURES
pn = input('Give test tunnel Project Name: ')
tu = input('Give test tunnel Tunnel Use. 1 for Sewer, 2 for Ventilation, 3 for Hydro, 4 for Rail, 5 for Train (lightrail), 6 for Metro, 7 for Road and 8 for Walk: ')
ct = input('Give test tunnel Cutting Technique. 1 for HAND, 2 for C&C, 3 for D&B, 4 for NATM/D&B,5 for NATM/OPEN, 6 for NATM, 7 for NATM/TBM, 8 for TBM/D&B, 9 for TBM/C&C and 10 for TBM: ')
l = input('Give test tunnel Length in meters: ')
d = input('Give test tunnel Diameter in meters: ')
o = input('Give test tunnel Overburden in meters: ')
rs = input('Give test tunnel Ground. 1 for Rock and 2 for Soil: ')

Give test tunnel Project Name: Test Tunnel Case
Give test tunnel Tunnel Use. 1 for Sewer, 2 for Ventilation, 3 for Hydro, 4 for Rail, 5 for Train (lightrail), 6 for Metro, 7 for Road and 8 for Walk: 7
Give test tunnel Cutting Technique. 1 for HAND, 2 for C&C, 3 for D&B, 4 for NATM/D&B,5 for NATM/OPEN, 6 for NATM, 7 for NATM/TBM, 8 for TBM/D&B, 9 for TBM/C&C and 10 for TBM: 3
Give test tunnel Length in meters: 6000
Give test tunnel Diameter in meters: 11
Give test tunnel Overburden in meters: 800
Give test tunnel Ground. 1 for Rock and 2 for Soil: 1
```

Figure 3. Input parameters for test tunnel case (TFM1).

```

The test tunnel can fall to the Very High Risk Category and bears a similarity of 89.12013840421298 % to the Very High Frequency Cluster's Centroid and a similarity of 99.76945744203302 % to the Nearest Neighbor Tunnel Failure Case.

The similarity of the test tunnel to the whole dataset is 69.71680751643456 %

The Nearest Neighbor Tunnel Failure Case is the following :
Project                Lambach Rail Tunnel
Tunnel_Use             Rail
Cutting_Technique      NATM
Length                 912
Diameter               12
Overburden             46
Rock_Soil              Rock
Cluster                1
Cluster_Label          Very High Frequency
Relative_Failure_Frequency_%  43.9252
Reshaped_Distance      0.00230543
Test_Tunnel_Similarity_to_Case_%  99.7695

The Relative Failure Probability of the test tunnel case is 39.14622901867299 %.

```

Figure 4. Results for the test tunnel case (TFM1).

At the last section of the project, the models' ability to predict the specific type of failure for a test tunnel was examined, by checking their prediction accuracy using failure cases from the dataset with a known result. For that reason, the dataset was separated in two sets, the training set which contains 90% or 80% of the incidents and the test set which contains 10% or 20%, respectively. The prediction is based on the failure type relative frequencies of the cluster to which the test tunnel can fall and on the failure type of the "nearest neighbor". The degree of influence of the "nearest neighbor" to the prediction is expressed by the factor W , which can take integer values from 1 to 5. The fact that W value can take values equal or greater to 1, indicates that for the developed methodology, the nearest neighbor's influence is considered as equal to, or stronger than the cluster's influence. The purpose of this part of the project was not only to evaluate the models' ability to predict, but also to determine the degree of influence of the nearest neighbor and the cluster's centroid, by finding the value of W that achieves the highest accuracy. The prediction accuracy of the models according to the W value and the separation ratio is shown in Table 4.

<i>Training/Test</i>	<i>Μοντέλο</i>	<i>Ακρίβεια (Accuracy) Μοντέλων (%)</i>				
		<i>W = 1</i>	<i>W = 2</i>	<i>W = 3</i>	<i>W = 4</i>	<i>W = 5</i>
90/10	TFM1 (K-Means)	0,42	0,75	0,75	0,75	0,75
	TFM2 (K-Prototypes)	0,67	0,75	0,92	0,92	0,92
80/20	TFM1 (K-Means)	0,40	0,70	0,70	0,70	0,70
	TFM2 (K-Prototypes)	0,45	0,45	0,60	0,60	0,60

Table 4. Prediction accuracy of the models according to the W value and the dataset separation ratio.

In conclusion, the performance of both models concerning the clustering function, the estimation of risk category, the calculation of relative failure probability and the calculation of similarity to the nearest neighbor and the cluster's centroid, is considered to be high and the results are considered as reliable. Model TFM2 presents a difficulty to manage data with extreme numerical parameter values and for this reason it is not recommended to use in such cases. As for the failure type prediction function, judging by the accuracy achieved, 75% and 70 for TFM1 and 92% and 60% for TFM2, the performance could be considered relatively high for both models. Although this research is actually a metadata analysis that has no primary data to assess, it is considered a significant effort for the classification and prediction of failures. Of course, the inclusion of primary data and geotechnical parameters could perhaps further enhance the accuracy as well as the consistency and reliability of the analysis. However, the increase of the accuracy due to the increase of the W value, suggests a strong influence of the "nearest neighbor" in the prediction making, that should be further researched.

Ευχαριστίες

Η παρούσα μεταπτυχιακή διπλωματική εργασία εκπονήθηκε στο πλαίσιο συμμετοχής μου στο Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών «Σχεδιασμός και Κατασκευή Υπόγειων Έργων», του Εθνικού Μετσόβιου Πολυτεχνείου.

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα της διπλωματικής μου εργασίας κ. Ανδρέα Μπενάρδο, Αναπληρωτή Καθηγητή της Σχολής Μηχανικών Μεταλλείων – Μεταλλουργών, για τη δυνατότητα που μου έδωσε να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα, τη συνεχή καθοδήγηση και υποστήριξη που μου παρείχε, καθώς και την εμπιστοσύνη που μου έδειξε.

Επιπλέον, θα ήθελα να ευχαριστήσω τον κ. Σπύρο Κωνσταντή, Διευθύνων Σύμβουλο της Ruler Consult Ltd, για την παροχή της βάσης δεδομένων, επί της οποίας εφαρμόζονται τα μοντέλα μηχανικής μάθησης που αναπτύχθηκαν προς εξόρυξη γνώσης για τον κίνδυνο αστοχίας σηράγγων.

Περιεχόμενα

Περίληψη.....	v
Abstract.....	vii
Ευχαριστίες.....	xiii
Περιεχόμενα	xv
Κατάλογος Εικόνων	xvii
Κατάλογος Σχημάτων.....	xxi
Κατάλογος Πινάκων.....	xxiii
Κατάλογος Διαγραμμάτων	xxv
1 Εισαγωγή.....	1
1.1 Παράθεση Προβλήματος	1
1.2 Υπό Εξέταση Αντικείμενο	1
1.3 Στόχος Εργασίας.....	2
1.4 Ακολουθούμενη Μεθοδολογία	3
1.5 Δομή Εργασίας	4
2 Κατασκευή Σηράγγων : Αβεβαιότητα, Αστοχίες - Απώλειες και Διαχείριση Κινδύνου.....	6
2.1 Αξιοπιστία - Αβεβαιότητα	7
2.2 Αστοχίες – Απώλειες	10
3 Μηχανική Μάθηση.....	13
3.1 Ορισμός Μηχανικής Μάθησης	13
3.2 Είδη Μηχανικής Μάθησης	13
3.3 Μη Επιβλεπόμενη Μάθηση	14
3.3.1 Συσταδοποίηση.....	14
3.3.2 Κατηγοριοποίηση Αλγορίθμων Συσταδοποίησης.....	16
3.3.3 Αλγόριθμος K-Means	17
3.3.4 Αλγόριθμος K-Prototypes.....	18
3.4 Εφαρμογές Μηχανικής Μάθησης για την Εκτίμηση Κινδύνου στην Κατασκευή Σηράγγων.....	20
4 Δεδομένα προς Ανάλυση.....	23
4.1 Περιγραφή Βάσης Δεδομένων.....	23
4.2 Προεπεξεργασία Δεδομένων	25
4.2.1 Επιλογή Χαρακτηριστικών.....	26

4.2.2	Πλήρωση Χαμένων Τιμών	28
4.2.3	Διαχείριση Θορυβωδών Δεδομένων.....	31
4.2.4	Μετατροπή Ονομαστικών Τιμών σε Αριθμητικές	31
4.2.5	Κανονικοποίηση Τιμών	32
4.3	Dataset Εφαρμογής Μοντέλων (TFD_21).....	33
4.4	Ανάλυση Κυρίων Συνιστωσών (Principal Component Analysis)	36
4.4.1	Περιγραφή Μεθόδου	36
4.4.2	Εφαρμογή Μεθόδου - Αποτελέσματα	40
5	Μοντέλα Ανάλυσης Δεδομένων.....	43
5.1	Μοντέλο TFM1.....	45
5.1.1	Συσταδοποίηση (TFM1_A).....	45
5.1.2	Εκτίμηση Κινδύνου Αστοχίας Σήραγγας (TFM1_B).....	49
5.2	Μοντέλο TFM2.....	54
5.2.1	Συσταδοποίηση (TFM2_A).....	54
5.2.2	Εκτίμηση Κινδύνου Αστοχίας Σήραγγας (TFM2_B).....	57
6	Αποτελέσματα Ανάλυσης Δεδομένων	60
6.1	Αποτελέσματα Μοντέλου TFM1	60
6.1.1	Αποτελέσματα Συσταδοποίησης (TFM1_A)	60
6.1.2	Αποτελέσματα Εκτίμησης Κινδύνου Αστοχίας Σήραγγας (TFM1_B)	65
6.2	Αποτελέσματα Μοντέλου TFM2.....	66
6.2.1	Αποτελέσματα Συσταδοποίησης (TFM2_A)	66
6.2.2	Αποτελέσματα Εκτίμησης Κινδύνου Αστοχίας Σήραγγας (TFM1_B)	70
7	Πρόβλεψη Τύπου Αστοχίας Σήραγγας (TFM1_C και TFM2_C).....	73
7.1	Μεθοδολογία Πρόβλεψης Τύπου Αστοχίας	73
7.2	Υπολογισμός Ακρίβειας Πρόβλεψης Τύπου Αστοχίας	78
8	Συμπεράσματα.....	81
8.1	Συμπεράσματα TFM1_A, TFM2_A.....	81
8.2	Συμπεράσματα TFM1_B, TFM2_B	85
8.3	Συμπεράσματα TFM1_C, TFM2_C	86
8.4	Προβλήματα – Προτάσεις Περαιτέρω Έρευνας.....	87
9	Βιβλιογραφία	89

Κατάλογος Εικόνων

<i>Εικόνα 1. Εισαγωγή dataset TFD_21 στο μοντέλο TFM1.</i>	45
<i>Εικόνα 2. Μετασχηματισμός περιγραφικών τιμών σε αριθμητικές.</i>	46
<i>Εικόνα 3. Κανονικοποίηση των τιμών με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ των τιμών 0 και 1.</i>	46
<i>Εικόνα 4. Εφαρμογή αλγορίθμου K-Means στο dataset TFD_21.</i>	47
<i>Εικόνα 5. Φθίνουσα κατάταξη σχετικής συχνότητας αστοχίας των cluster (TFM1).</i>	48
<i>Εικόνα 6. Εισαγωγή των στηλών Cluster_Label και Relative_Failure_Frequency_% στο dataset (TFM1).</i>	48
<i>Εικόνα 7. Εισαγωγή από το χρήστη των χαρακτηριστικών της υπό εξέταση σήραγγας (TFM1).</i>	49
<i>Εικόνα 8. Προσθήκη της υπό εξέταση σήραγγας (Test Tunnel Case) ως εγγραφή (γραμμή 107) στο dataset (TFM1).</i>	50
<i>Εικόνα 9. Κανονικοποίηση των τιμών, συμπεριλαμβανομένων και αυτών που αφορούν στην υπό εξέταση σήραγγα, με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ των τιμών 0 και 1.</i>	50
<i>Εικόνα 10. Υπολογισμός ευκλείδειας απόστασης υπό εξέταση σήραγγας από τα κεντροειδή των συστάδων (TFM1).</i>	52
<i>Εικόνα 11. Αναγωγή (μετασχηματισμένης) απόστασης από το κεντροειδές σε ποσοστιαία ομοιότητα με το κεντροειδές (Similarity to Cluster's Centroid %) (TFM1).</i>	52
<i>Εικόνα 12. Υπολογισμός ομοιότητας της υπό εξέταση σήραγγας με κάθε περιστατικό του dataset και εισαγωγή της αντίστοιχης στήλης (Test Tunnel Similarity to Case %) (TFM1).</i>	52
<i>Εικόνα 13. Εισαγωγή dataset TFD_21 στο μοντέλο TFM2.</i>	54
<i>Εικόνα 14. Κανονικοποίηση των αριθμητικών τιμών με τη μέθοδο ελάχιστου–μέγιστου, μεταξύ 0 και 1.</i>	55
<i>Εικόνα 15. Εφαρμογή αλγορίθμου K-Prototypes στο dataset TFD_21.</i>	55
<i>Εικόνα 16. Φθίνουσα κατάταξη σχετικής συχνότητας αστοχίας των clusters (TFM2).</i>	56
<i>Εικόνα 17. Εισαγωγή των στηλών Cluster_Label και Relative Failure Frequency % στο dataset (TFM2).</i>	56

<i>Εικόνα 18. Εισαγωγή από το χρήστη των χαρακτηριστικών της υπό εξέταση σήραγγας (TFM2).</i>	57
<i>Εικόνα 19. Προσθήκη της υπό εξέταση σήραγγας (Test Tunnel Case) ως εγγραφή (γραμμή 107) dataset (TFM2).</i>	57
<i>Εικόνα 20. Κανονικοποίηση των αριθμητικών τιμών, συμπεριλαμβανομένων και αυτών που αφορούν στην υπό εξέταση σήραγγα, με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ των τιμών 0 και 1.</i>	58
<i>Εικόνα 21. Υπολογισμός “απόστασης” υπό εξέταση σήραγγας από τα κεντροειδή των συστάδων.</i>	59
<i>Εικόνα 22. Αναγωγή “απόστασης” από το κεντροειδές σε ποσοστιαία ομοιότητα με το κεντροειδές (Similarity to Cluster’s Centroid %).</i>	59
<i>Εικόνα 23. Υπολογισμός ομοιότητας της υπό εξέταση σήραγγας με κάθε περιστατικό του dataset και εισαγωγή της αντίστοιχης στήλης (Test Tunnel Similarity to Case %) (TFM2).</i>	59
<i>Εικόνα 24. Ποσοστιαία εμφάνιση τιμών του περιγραφικού χαρακτηριστικού tunnel use ανά cluster (TFM1).</i>	62
<i>Εικόνα 25. Διαγραμματική απεικόνιση χαρακτηριστικών cutting technique και rock/soil, ανά cluster (TFM1).</i>	64
<i>Εικόνα 26. Εισαγωγή χαρακτηριστικών υπό εξέταση σήραγγας στο μοντέλο TFM1.</i>	65
<i>Εικόνα 27. Αποτελέσματα εφαρμογής του μοντέλου TFM1 στην υπό εξέταση σήραγγα.</i>	65
<i>Εικόνα 28. Ποσοστιαία εμφάνιση τιμών του περιγραφικού χαρακτηριστικού tunnel use ανά cluster (TFM2).</i>	68
<i>Εικόνα 29. Διαγραμματική απεικόνιση χαρακτηριστικών cutting technique και rock/soil, ανά cluster (TFM2).</i>	70
<i>Εικόνα 30. Εισαγωγή χαρακτηριστικών υπό εξέταση σήραγγας στο μοντέλο TFM2.</i>	71
<i>Εικόνα 31. Αποτελέσματα εφαρμογής του μοντέλου TFM2 στην υπό εξέταση σήραγγα.</i>	71
<i>Εικόνα 32 Σχετική συχνότητα εμφάνισης τύπων αστοχίας (στήλη Failure_Type_Frequency) που εμφανίζονται στο cluster στο οποίο εμπίπτει μια προς εξέταση σήραγγα.</i>	74
<i>Εικόνα 33. Στοιχεία "κοντινότερου γείτονα" της υπό εξέταση σήραγγας (τύπος αστοχίας OSF: Overstressing Support Failure).</i>	75

<i>Εικόνα 34. Αποτελέσματα εφαρμογής μοντέλου για τον υπολογισμό της σχετικής πιθανότητας εμφάνισης συγκεκριμένου τύπου αστοχίας για μία υπό εξέταση σήραγγα.....</i>	<i>77</i>
<i>Εικόνα 35. Αποτελέσματα πρόβλεψης τύπου αστοχίας σήραγγας για τα μοντέλα TFM1 και TFM2 ($W = 4$, αναλογία διαχωρισμού 90/10).</i>	<i>80</i>
<i>Εικόνα 36. Διαγραμματική απεικόνιση χαρακτηριστικών cutting technique και rock soil ανά cluster (αριστερά TFM1, δεξιά TFM2).</i>	<i>82</i>

Κατάλογος Σχημάτων

Σχήμα 1. Πίνακας αξιολογίας γεωλογικού μοντέλου (GMR, Geological Model Rating) [7].	7
Σχήμα 2. Πίνακας προσδιορισμού πολυπλοκότητας γεωλογικού μοντέλου και αβεβαιότητας μετά από ορισμένη έκταση έρευνας πεδίου [10]......	9
Σχήμα 3. Διάγραμμα προσδιορισμού ακρίβειας γεωτεχνικού μοντέλου [2] [3].	9
Σχήμα 4. Κατανομή οικονομικών απωλειών ασφαλιστικής αγοράς για την κάλυψη έργων κατασκευής σηράγγων [5]......	10
Σχήμα 5. Κατανομή χρονικών καθυστερήσεων έργων κατασκευής σηράγγων [5].	11
Σχήμα 6. Συσχέτιση οικονομικών απωλειών και χρονικής καθυστέρησης έργων [5].	11
Σχήμα 7. Οικονομικές απώλειες ανά τύπο αστοχίας [5].	12
Σχήμα 8. Οικονομικές απώλειες ανά μέθοδο κατασκευής [5]......	12
Σχήμα 9. Βήματα της διαδικασίας συσταδοποίησης [13].	15
Σχήμα 10. Γεωυλικό εντός του οποίου κατασκευάστηκαν οι σήραγγες που περιλαμβάνονται: α. στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων και β. στο dataset TFD_21.	41
Σχήμα 11. Μέθοδος κατασκευής των σηράγγων που περιλαμβάνονται: α. στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων και β. στο dataset TFD_21.	41
Σχήμα 12. Γεωυλικό σε συσχέτιση με τη μέθοδο κατασκευής των σηράγγων που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων (αριστερά) και στο dataset TFD_21 (δεξιά).	41
Σχήμα 13. Απεικόνιση μέσω εφαρμογής PCA, σε δύο διαστάσεις, υπό εξέταση σήραγγας (test tunnel) και εγγύτητας αυτής με τα κεντροειδή των συστάδων (Euclidian distance $e.d.1. < e.d.0 < e.d.3 < e.d.2$).	44

Κατάλογος Πινάκων

Πίνακας 1. Στοιχεία - χαρακτηριστικά περιστατικών αστοχίας σηράγγων που συμπεριλαμβάνονται στη βάση δεδομένων και αντίστοιχα ποσοστά πλήρωσης.	24
Πίνακας 2. Στοιχεία - χαρακτηριστικά περιστατικών αστοχίας σηράγγων που συμπεριλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων και αντίστοιχα ποσοστά πλήρωσης.....	27
Πίνακας 3. Χαρακτηριστικά περιστατικών αστοχίας σηράγγων που συμπεριλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων, μετά την πλήρωση των χαμένων τιμών, και αντίστοιχα ποσοστά πλήρωσης.....	28
Πίνακας 4. Μετατροπή ονομαστικών - περιγραφικών τιμών σε αριθμητικές (για την εφαρμογή του αλγόριθμου K-Means).	32
Πίνακας 5. Σχετικές συχνότητες αστοχίας και τα αντίστοιχα labels των clusters που προέκυψαν από τη εφαρμογή του TFM1 στο dataset.	60
Πίνακας 6. Ποσοστιαία αναλογία εμφάνισης τιμών που αφορούν στα περιγραφικά χαρακτηριστικά των σηράγγων, ανά cluster (TFM1)	61
Πίνακας 7.Πρότυπα (μοτίβα) χαρακτηριστικών που προκύπτουν από την διαμόρφωση των cluster (TFM1).....	64
Πίνακας 8. Σχετικές συχνότητες αστοχίας και τα αντίστοιχα labels των clusters που προέκυψαν από τη εφαρμογή του TFM2 στο dataset.	66
Πίνακας 9. Ποσοστιαία αναλογία εμφάνισης τιμών που αφορούν στα περιγραφικά χαρακτηριστικά των σηράγγων, ανά cluster (TFM1)	67
Πίνακας 10. Πρότυπα (μοτίβα) χαρακτηριστικών που προκύπτουν από τη διαμόρφωση των cluster (TFM2).....	70
Πίνακας 11. Τιμές χαρακτηριστικού τύπος αστοχίας (failure type) και αντίστοιχες συντομογραφίες	73
Πίνακας 12. Ακρίβεια (accuracy) των μοντέλων TFM1 και TFM2, ως προς την πρόβλεψη τύπου αστοχίας σήραγγας, για τις διάφορες τιμές του συντελεστή βαρύτητας W	79
Πίνακας 13. Σχετική συχνότητα (relative frequency) εμφάνισης των cluster που προέκυψαν ανά μοντέλο και ποσοστιαία συμφωνία εγγραφών.	81

<i>Πίνακας 14. Πρότυπα (μοτίβα) χαρακτηριστικών που προκύπτουν από την διαμόρφωση των cluster.</i>	<i>83</i>
<i>Πίνακας 15. Πρότυπα (μοτίβα) χαρακτηριστικών που συσχετίζονται με ορισμένη κατηγορία κινδύνου (μέσω αξιολόγησης - συνδυασμού προτύπων TFM1 και TFM2).....</i>	<i>84</i>

Κατάλογος Διαγραμμάτων

Διάγραμμα 1. Γεωυλικό εντός του οποίου κατασκευάστηκαν οι σήραγγες που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων.....	29
Διάγραμμα 2. Μέθοδος κατασκευής σιηράγγων που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων.....	29
Διάγραμμα 3. Γεωυλικό σε συσχέτιση με τη μέθοδο κατασκευής των σιηράγγων που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων.....	29
Διάγραμμα 4. Τύποι αστοχίας των σιηράγγων που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων (FAC: Face Collapse, OSF: Overstressing/Support Failure, FIR: Fire, CRW: Crown Wedge Failure, LS: Lateral Support Failure, OTH: Other).	30
Διάγραμμα 5. Γεωυλικό εντός του οποίου κατασκευάστηκαν οι σήραγγες που περιλαμβάνονται στο dataset TFD_21.....	34
Διάγραμμα 6. Μέθοδος κατασκευής των σιηράγγων που περιλαμβάνονται στο dataset TFD_21.....	34
Διάγραμμα 7. Γεωυλικό σε συσχέτιση με τη μέθοδο κατασκευής των σιηράγγων που περιλαμβάνονται στο dataset TFD_21.....	35
Διάγραμμα 8. Τύποι αστοχίας των σιηράγγων που περιλαμβάνονται στο dataset TFD_21 (FAC: Face Collapse, OSF: Overstressing/Support Failure, FIR: Fire, CRW: Crown Wedge Failure, LS: Lateral Support Failure, OTH: Other).	35
Διάγραμμα 9. Απεικόνιση των δεδομένων που περιλαμβάνονται στο dataset TFD_21 σε δύο διαστάσεις, ως προς τις κύριες συνιστώσες $pc1$ και $pc2$, που προέκυψαν από την εφαρμογή της PCA. Εμφανής σχηματισμός συστάδων (clusters).....	42
Διάγραμμα 10. Μείωση διαστατικότητας μέσω εφαρμογής PCA και απεικόνιση των δεδομένων υπό τις δύο κύριες συνιστώσες ($pc1$ και $pc2$).....	47

1 Εισαγωγή

1.1 Παράθεση Προβλήματος

Η κατασκευή υπόγειων έργων και ειδικότερα η κατασκευή σιράγγων, σε αντίθεση με την κατασκευή κτιριακών έργων ή άλλων δομημάτων ανωδομής, υλοποιείται σε ένα περιβάλλον όπου η γνώση των επικρατούσων συνθηκών χαρακτηρίζεται από μεγάλη αβεβαιότητα. Παρά το γεγονός ότι οι φορείς υποστήριξης που χρησιμοποιούνται είναι από σκυρόδεμα και χάλυβα, των οποίων οι ιδιότητες είναι γνωστές, ο κύριος φέρων οργανισμός είναι η ίδια η γεώμαζα, η γνώση των ιδιοτήτων της οποίας στις περισσότερες περιπτώσεις είναι περιορισμένη [1]. Η επιτυχής υλοποίηση του έργου ή η ύπαρξη αστοχιών κατά την κατασκευή, είναι κατά κύριο λόγο αποτέλεσμα της ακρίβειας των εκτιμήσεων και των υπολογισμών που πραγματοποιήθηκαν κατά την εκπόνηση της μελέτης του έργου, λαμβάνοντας υπόψη το γεωλογικό μοντέλο [2]. Δυστυχώς όμως, η πληροφορία που δύναται να συγκεντρωθεί για την ανάπτυξη του γεωλογικού μοντέλου, μέσω εφαρμογής όλων των δυνατών μεθόδων έρευνας του υπεδάφους (γεωτεχνικών και γεωφυσικών), είναι περιορισμένη, καθότι οι εργασίες που πρέπει να πραγματοποιηθούν εξαρτώνται από τους φυσικούς περιορισμούς που επιβάλλει το ίδιο το περιβάλλον του έργου, αλλά και από οικονομικούς παράγοντες. Το ποσοστό του προϋπολογισμού του έργου ή το ποσό που προτίθεται η εκάστοτε κατασκευάστρια εταιρία να διαθέσει για τις εργασίες αυτές, είναι ορισμένο και ως εκ τούτου ο βαθμός βεβαιότητας ισχύος του παραγόμενου γεωλογικού μοντέλου είναι συγκεκριμένος. Η δυνατότητα εκτίμησης της πιθανότητας αστοχίας, λαμβάνοντας όμως υπόψη συγκεκριμένα χαρακτηριστικά του έργου τα οποία είναι γνωστά εκ των προτέρων, επιτρέπει να ορίσουμε και την ελάχιστη απαιτούμενη ακρίβεια του γεωλογικού μοντέλου. Επιπλέον, γνωρίζοντας και την επίδραση μιας επικείμενης αστοχίας, μπορούμε να εκτιμήσουμε και τον αντίστοιχο κίνδυνο, υπό την έννοια απρόβλεπτο κόστους αλλά και ημερολογιακής καθυστέρησης.

1.2 Υπό Εξέταση Αντικείμενο

Αντικείμενο της παρούσας εργασίας, είναι η ανάπτυξη και εφαρμογή μοντέλων μηχανικής μάθησης για την ομαδοποίηση περιστατικών αστοχίας σιράγγων ανά τον κόσμο, αλλά και η προσπάθεια αξιοποίησης της εμπειρίας ώστε να μπορέσουμε να αναγνωρίσουμε

πρότυπα και συμπεριφορές που πιθανώς θα οδηγήσουν σε νέες τέτοιες συνθήκες σε μελλοντικά έργα. Η ομαδοποίηση πραγματοποιείται με τη χρήση αλγορίθμων συσταδοποίησης (clustering) που ενσωματώνονται στα μοντέλα, βάσει συσχετίσεων και προτύπων (μοτίβων) που παρατηρούνται στα κύρια χαρακτηριστικά των έργων, όπως η διάμετρος, η μέθοδος διάνοιξης, το πάχος των υπερκείμενων κ.α. Τα περιστατικά αυτά περιλαμβάνονται σε μία βάση δεδομένων, η οποία αποτελεί προϊόν πρότερων ερευνητικών εργασιών αντίστοιχου αντικειμένου. Αρχικά πραγματοποιείται μία προεπεξεργασία των δεδομένων, μέσω της οποίας ορίζεται ένα εκμεταλλεύσιμο τμήμα της βάσης δεδομένων, επί του οποίου στη συνέχεια εφαρμόζονται μοντέλα. Η εξόρυξη γνώσης σχετικά με την αστοχία σήραγγων, πραγματοποιείται μέσω της ερμηνείας των αποτελεσμάτων εφαρμογής των μοντέλων, ήτοι η διαμόρφωση ομάδων περιστατικών, μέσω αναγνώρισης των πιθανών προτύπων (μοτίβων) χαρακτηριστικών που απαντώνται σε αυτές. Στην παρούσα εργασία εξετάζεται η εφαρμογή δύο μοντέλων με διαφορετική αρχιτεκτονική, των οποίων η αποτελεσματικότητα αξιολογείται, με σκοπό την εύρεση του καταλληλότερου μοντέλου για την απαιτούμενη χρήση.

1.3 Στόχος Εργασίας

Στόχος της εργασίας είναι ο σχεδιασμός ενός κατάλληλου μοντέλου μηχανικής μάθησης, μέσω του οποίου αρχικά θα πραγματοποιείται η ομαδοποίηση των συμβάντων αστοχίας που περιλαμβάνονται στη βάση δεδομένων και θα ερμηνεύονται τα αποτελέσματα, προς τυχόν παρατήρηση προτύπων (μοτίβων) που οδηγούν σε συγκεκριμένο βαθμό κινδύνου αστοχίας. Στη συνέχεια εισάγοντας τις παραμέτρους μίας νέας υπό εξέταση σήραγγας, θα προκύπτει η εκτιμώμενη σχετική πιθανότητα αστοχίας και ο βαθμός ομοιότητας της υπό εξέταση σήραγγας με άλλα έργα της βάσης δεδομένων. Η πιθανότητα αστοχίας προκύπτει μέσω σύγκρισης της υπό εξέταση σήραγγας, με εκείνες που περιλαμβάνονται στη βάση δεδομένων που ήταν διαθέσιμη και αφορά στην πιθανότητα εμφάνισης αστοχίας γενικά, αλλά και στην πιθανότητα εμφάνιση αστοχίας συγκεκριμένου τύπου. Δεδομένου ότι είναι γνωστά ο τύπος, τα χαρακτηριστικά και τα αποτελέσματα της αστοχίας, υπάρχει η δυνατότητα εξαγωγής χρήσιμων συμπερασμάτων για τον κίνδυνο αστοχίας που αντιμετωπίζει η εκάστοτε υπό εξέταση σήραγγα. Απώτερος στόχος είναι η ενσωμάτωση του κώδικα σε μία εφαρμογή H/Y, η οποία θα αποτελέσει χρήσιμο εργαλείο για τις ασφαλιστικές εταιρίες αλλά και τις υποψήφιες αναδόχους κατασκευάστριες εταιρίες

έργων σηράγγων. Είναι προφανές ότι σε περίπτωση που θα ήταν δυνατή η ενσωμάτωση περαιτέρω στοιχείων και δεδομένων από περισσότερες σήραγγες, τότε θα ήταν δυνατή η βελτίωση της εφαρμοσιμότητας αλλά και ακρίβειας πρόβλεψης του μοντέλου.

1.4 Ακολουθούμενη Μεθοδολογία

Στο πρώτο στάδιο της εργασίας πραγματοποιείται η προεπεξεργασία των δεδομένων, οι εργασίες δηλαδή προετοιμασίας των δεδομένων που περιέχονται στην βάση πριν την καθαυτή εξόρυξη της γνώσης. Αρχικά γίνεται ο “καθαρισμός” των δεδομένων (εύρεση - διόρθωση αλληλοσυγκρουόμενων πληροφοριών, πολλαπλών εγγραφών κ.α.), στη συνέχεια επιλέγονται τα *γνωρίσματα - χαρακτηριστικά* που θα χρησιμοποιηθούν για την εφαρμογή των αλγορίθμων και τέλος συμπληρώνεται κατά το δυνατόν η βάση δεδομένων, με πλήρωση μόνο των απαιτούμενων για την εφαρμογή των αλγορίθμων χαρακτηριστικών.

Το δεύτερο στάδιο αποτελείται από το πρώτο και το δεύτερο μέρος σχεδιασμού και εφαρμογής των μοντέλων. Το πρώτο μέρος αφορά στην *ομαδοποίηση - συσταδοποίηση* των εγγραφών (δηλαδή των περιστατικών αστοχίας σηράγγων) της βάσης δεδομένων και στην ερμηνεία των αποτελεσμάτων της, ενώ το δεύτερο μέρος αφορά στην εκτίμηση του κινδύνου αστοχίας μίας νέας υπό εξέταση σήραγγας, της οποίας εισάγονται οι παράμετροι στο εκάστοτε μοντέλο. Το πρώτο μοντέλο (TFM1) βασίζεται στον αλγόριθμο συσταδοποίησης (*clustering*) *K-Means* και το δεύτερο μοντέλο (TFM2) στον αλγόριθμο συσταδοποίησης *K-Prototypes*.

Στο τρίτο στάδιο της εργασίας τα δύο μοντέλα “εκπαιδεύονται” με χρήση συγκεκριμένου τμήματος της βάσης δεδομένων (*training set*) και υπολογίζεται η ακρίβεια πρόβλεψης συγκεκριμένου τύπου αστοχίας για τις υπόλοιπες εγγραφές της βάσης (*test set*), προκειμένου να εξαχθούν συμπεράσματα που αφορούν στην αποτελεσματικότητα των δύο (2) μοντέλων και να αξιολογηθούν συγκριτικά οι δύο αλγόριθμοι, ως προς την καταλληλότητα χρήσης για το εξεταζόμενο αντικείμενο.

1.5 Δομή Εργασίας

Το 1^ο Κεφάλαιο αποτελεί την εισαγωγή της εργασίας, όπου παρατίθεται το πρόβλημα και παρουσιάζεται το υπό εξέταση αντικείμενο και η σημασία του, ο στόχος της εργασίας, η ακολουθούμενη μεθοδολογία και η δομή της εργασίας.

Στο 2^ο Κεφάλαιο αναλύεται η αβεβαιότητα που χαρακτηρίζει τα έργα κατασκευής σηράγγων και οι απώλειες που επιφέρουν οι απρόβλεπτες αστοχίες που συνεπάγονται αυτής της αβεβαιότητας. Επιπλέον, παρατίθενται ορισμένα ποιοτικά και ποσοτικά στοιχεία που προκύπτουν από την διεθνή βιβλιογραφία και αφορούν στη συσχέτιση των χαρακτηριστικών των έργων με τις προκληθείσες απώλειες καθώς και οι διεθνείς πρακτικές διαχείρισης του κινδύνου στον οποίο υπόκεινται τα έργα κατασκευής σηράγγων.

Στο 3^ο Κεφάλαιο παρατίθεται το θεωρητικό υπόβαθρο του υπό διερεύνηση αντικειμένου, καθορίζονται οι βασικοί όροι και έννοιες που παρουσιάζονται στην συνέχεια της εργασίας και περιγράφονται συνοπτικά οι αλγόριθμοι και οι τεχνικές στις οποίες βασίστηκε η δημιουργία των μοντέλων. Επιπλέον αναφέρονται ορισμένες εφαρμογές της μηχανικής μάθησης για τη διαχείριση κινδύνου στην κατασκευή σηράγγων.

Στο 4^ο Κεφάλαιο παρουσιάζεται η βάση δεδομένων και τα στατιστικά στοιχεία τα οποία προκύπτουν από την ανάλυση αυτής και αναλύεται ως προς το ποσοστό πλήρωσής της. Στη συνέχεια περιγράφεται η προεπεξεργασία που πραγματοποιήθηκε στα δεδομένα, προκειμένου να προκύψουν εκείνα τα οποία είναι δυνατόν να ληφθούν υπόψη για τη διαδικασία συσταδοποίησης και παρουσιάζεται το τελικό τμήμα του συνόλου δεδομένου επί του οποίου εφαρμόστηκαν τα μοντέλα - αλγόριθμοι.

Στο 5^ο Κεφάλαιο αρχικά αναλύονται τα διακριτά τμήματα των μοντέλων που αφορούν στη διαδικασία συσταδοποίησης των δεδομένων, με αναφορά στην επιτελεσματικότητα της κάθε χρησιμοποιούμενης τεχνικής και στον τρόπο ενσωμάτωσης των αλγορίθμων και εφαρμογής των τεχνικών, υπό μορφή κώδικα. Στη συνέχεια αναλύεται η μεθοδολογία εκτίμησης κινδύνου αστοχίας μίας νέας υπό εξέταση σήραγγας, με παράλληλη παράθεση των κωδικοποιημένων τμημάτων του κάθε μοντέλου.

Στο 6^ο Κεφάλαιο παρουσιάζονται τα αποτελέσματα της συσταδοποίησης που προέκυψαν από την εφαρμογή των δύο μοντέλων και σχολιάζονται - αξιολογούνται οι διαφορές στη δομή των συστάδων προς εκτίμηση πιθανών προτύπων. Επιπλέον,

παρουσιάζεται η εφαρμογή των δύο μοντέλων, για την εκτίμηση της σχετικής πιθανότητας αστοχίας μίας υπό εξέταση σήραγγας.

Στο 7^ο Κεφάλαιο αρχικά αναλύεται η μέθοδος υπολογισμού ακρίβειας πρόβλεψης τύπου αστοχίας και στη συνέχεια πραγματοποιείται υπολογισμός ακρίβειας πρόβλεψης του κάθε μοντέλου. Σκοπός υλοποίησης αυτού του σκέλους της εργασίας, είναι η αξιολόγηση της δυνατότητας πρόβλεψης των μοντέλων, η εύρεση των παραγόντων που την επηρεάζουν, αλλά και η συγκριτική αξιολόγηση των μοντέλων.

Στο 8^ο Κεφάλαιο συνοψίζεται η εργασία, αναπτύσσονται τα κύρια ευρήματα – συμπεράσματα και παρατίθενται προτάσεις για την περαιτέρω έρευνα του αντικειμένου.

2 Κατασκευή Σηράγγων : Αβεβαιότητα, Αστοχίες - Απώλειες και Διαχείριση Κινδύνου

Οι παραδοσιακές μέθοδοι υπολογισμού κόστους, οι οποίες βασίζονται στην αξιοποίηση στοιχείων κόστους παλαιότερων έργων, δεν δύναται να εφαρμοστούν στην περίπτωση των έργων κατασκευής σηράγγων και γενικότερα στην κατασκευή υπόγειων έργων. Το κόστος σε αυτή την περίπτωση διαφοροποιείται σημαντικά λόγω της ποικιλομορφίας του υπόγειου περιβάλλοντος αλλά και λόγω της εξάρτησής του από τη γεωγραφία και τις ειδικές παραμέτρους του εκάστοτε έργου [3]. Επιπλέον, η αδυναμία ακριβούς προσδιορισμού των γεωτεχνικών παραμέτρων, λόγω των περιορισμών που επιβάλλει το ίδιο το περιβάλλον, έχει ως αποτέλεσμα την δημιουργία γεωλογικών μοντέλων με μεγάλο βαθμό αβεβαιότητας. Η εγγενής αυτή αβεβαιότητα, εκφράζεται πολλές φορές μέσω απρόβλεπτων γεγονότων που επιφέρουν σημαντικές οικονομικές απώλειες, υπερβάσεις προϋπολογισμών και χρονικές καθυστερήσεις στα έργα. Οι χρονικές καθυστερήσεις όμως για τα μεγάλα έργα υποδομής μεταφορών, στα οποία εμπίπτουν κατά κύριο λόγο οι κατασκευές σηράγγων, είναι ο μεγαλύτερος παράγοντας ανησυχίας που μπορεί να απειλήσει την βιωσιμότητα του έργου και κατά συνέπεια, για την επιτυχή ολοκλήρωση ενός τέτοιου έργου είναι απαραίτητος ο σχεδιασμός και η οργάνωση ενός πλαισίου διαχείρισης κινδύνου, το οποίο θα περιλαμβάνει τον εντοπισμό, την αξιολόγηση και την ελαχιστοποίηση των κινδύνων, πριν την απόφαση υλοποίησης του έργου και κατά τη διάρκεια εκτέλεσής του [4]. Στο πλαίσιο αυτό, της διαχείρισης του κινδύνου, εμπλέκονται οι κατασκευαστές των έργων αλλά και οι ασφαλιστικές εταιρίες οι οποίες αναλαμβάνουν να καλύψουν μέρος της οικονομικής απώλειας [5].

Η αυξημένη πιθανότητα πρόκλησης απρόβλεπτων περιστατικών κυρίως κατά τη φάση κατασκευής των σηράγγων, αλλά και οι σημαντικές απώλειες που επιφέρουν αυτά τα περιστατικά, οδήγησαν στην ανάγκη ανάπτυξης μεθόδων διαχείρισης κινδύνου για την διασφάλιση τόσο των έργων όσο και των εμπλεκόμενων σε αυτά. Οι σημαντικότερες ενέργειες μέχρι στιγμής σε αυτόν το τομέα, που είχαν ως σκοπό τον μετριασμό των δυνητικών μελλοντικών απωλειών, ήταν η εισαγωγή της Έκθεσης Τιμών Αναφοράς (Geotechnical Baseline Report, GBR) και η εισαγωγή του Κώδικα Πρακτικής για τη Διαχείριση του Κινδύνου στα Έργα Σηράγγων (Code of Practice for Risk Management of Tunnel Works) [6].

2.1 Αξιοπιστία - Αβεβαιότητα

Η κατασκευή σηράγγων εμπεριέχει μεγάλο βαθμό αβεβαιότητας, η οποία αφορά κυρίως στη συμπεριφορά του περιβάλλοντος γεωολογικού κατά τη διάρκεια της διάνοιξης και της τοποθέτησης της υποστήριξης. Η πολυπλοκότητα του γεωολογικού μοντέλου, εξαρτάται από ένα πλήθος παραγόντων όπως η γεωολογική ιστορία, η τεκτονική, η γεωδυναμική, η υδρογεωλογία κ.α.. Η πολυπλοκότητα όμως του γεωολογικού μοντέλου σε συνδυασμό με την έκταση της έρευνας πεδίου μέσω των διαθέσιμων γεωτεχνικών και γεωφυσικών μεθόδων, είναι οι παράγοντες που εν τέλει ορίζουν τον βαθμό βεβαιότητας ισχύος του παραγόμενου γεωτεχνικού μοντέλου και ως εκ τούτου την πιθανότητα αστοχίας της σήραγγας και τον κίνδυνο εκδήλωσής της.

Η αξιοπιστία του γεωολογικού μοντέλου, μελετήθηκε το 2011 από τον Perello [7], ο οποίος ανέπτυξε μία υπολογιστική μεθοδολογία εκτίμησης αυτής. Για την εφαρμογή της μεθοδολογίας, η σήραγγα διαχωρίζεται σε διακριτά ισομήκη τμήματα (συνηθώς 100 m) και για κάθε ένα από αυτά προσδιορίζονται οι τιμές διαφόρων παραμέτρων, υπό τη μορφή “check list”. Μέσω μίας υπολογιστικής διαδικασίας προκύπτει για κάθε τμήμα της σήραγγας μία τιμή GMR (Geological Model Rating), η οποία αφορά στην αξιοπιστία του γεωολογικού μοντέλου, για κάθε διακριτό τμήμα της σήραγγας. Οι τιμές του GMR κυμαίνονται από 0 έως 10 και η σημασία του κάθε εύρους τιμών φαίνεται στο Σχήμα 1.

		a	b	c	d	e
GMR = 10 – 7.5 →	Good reliability	No	Possible <i>Imprecision order of magnitude 0-50m</i>	Possible <i>Imprecision order of magnitude 0-25%</i>	No	No
GMR = 7.5 – 5 →	Fair reliability	No	Possible <i>Imprecision order of magnitude 0-100m</i>	Possible <i>Imprecision order of magnitude 0-50%</i>	Possible <i>but not probable</i>	No
GMR = 5 – 2.5 →	Poor reliability	Possible <i>but not probable</i>	Possible <i>Imprecision order of magnitude 0-200m</i>	Possible <i>Imprecision order of magnitude 0-100%</i>	Possible	Possible <i>but not probable</i>
GMR = 2.5 – 0 →	Unreliable	Possible	Possible <i>Imprecision order of magnitude >200m</i>	Possible <i>Imprecision order of magnitude >100%</i>	Possible	Possible

a = Existence of significant deviations with regard to the RGM;
b = Imprecision in the position of lithological or fault zones contacts;
c = Imprecision in the thickness of lithological levels or fault zones;
d = Presence of further critical geological elements of secondary importance besides the forecasted ones (metric to decametric faults/levels with poor geomechanical conditions);
e = Presence of further critical geological elements of primary importance besides the forecasted ones (decametric to pluri-decametric faults/levels with poor geomechanical conditions)

Σχήμα 1. Πίνακας αξιοπιστίας γεωολογικού μοντέλου (GMR, Geological Model Rating) [7].

Η αβεβαιότητα στα γεωλογικά συστήματα χωρίζεται στις δύο παρακάτω κατηγορίες [8]:

α. Την επιστημική αβεβαιότητα, η οποία οφείλεται στις φυσικές διαδικασίες που δεν μπορούν να προσδιοριστούν με λεπτομέρεια, δηλαδή στη συμπεριφορά της γης που επηρεάζει ολόκληρη την πιθανολογική ανάλυση [9].

β. Την στοχαστική αβεβαιότητα, η οποία οφείλεται στις φυσικές διαδικασίες που δεν μπορούν να προσδιοριστούν με λεπτομέρεια, δηλαδή στη συμπεριφορά της γης που επηρεάζει ολόκληρη την πιθανολογική ανάλυση. Αντιπροσωπεύει λεπτομέρειες την απόκρισης της πηγής, του δρόμου διάδοσης και του σημείου ενδιαφέροντος που δεν μπορούν να ποσοτικοποιηθούν πριν συμβεί κάποιος σεισμός. Ακόμη και να εξαλειφόταν τελείως η επιστημική αβεβαιότητα, η κυβευτική αβεβαιότητα θα εξακολουθούσε να υπάρχει, και φυσικά θα επηρέαζε τα αποτελέσματα [9].

Οι ανωτέρω παράγοντες (κατηγορίες αβεβαιότητας) περιορίζουν την ικανότητα ανάλυσης και μοντελοποίησης του ρεαλιστικού προβλήματος. Ο όρος «αβέβαιο» αναφέρεται στις τιμές που δεν είναι προβλέψιμες και μπορεί να είναι άγνωστες. Η αβεβαιότητα των γεωλογικών και γεωτεχνικών προβλημάτων μπορεί να χωριστεί στις παρακάτω τρεις κύριες κατηγορίες [8]:

α. Αβεβαιότητα χαρακτηρισμού τοποθεσίας: αναφέρεται στην ακρίβεια του γεωλογικού μοντέλου, το οποίο επηρεάζεται από την αβεβαιότητα των δεδομένων και της έρευνας πεδίου.

β. Αβεβαιότητα παραμέτρων: προκύπτει από την ανακρίβεια αξιολόγησης των γεωτεχνικών παραμέτρων που προκύπτουν από τις δοκιμές.

γ. Αβεβαιότητα μοντέλου: αφορά στον βαθμό κατά τον οποίο το μαθηματικό μοντέλο δύναται να αποδώσει επαρκώς την πραγματικότητα.

Η εν λόγω αβεβαιότητα δύναται να οδηγήσει σε υπέρβαση του αρχικού προϋπολογισμού των έργων και σε σημαντικές καθυστερήσεις, οι οποίες οφείλονται κατά κύριο λόγο σε ανεπιτυχή έρευνα του πεδίου (υπεδάφους) στο οποίο θα πραγματοποιηθεί η διάνοιξη της σήραγγας [8].

Το 2019 αναπτύχθηκε από τους Venturini et al. [10] μία μεθοδολογία εκτίμησης της αβεβαιότητας του γεωτεχνικού μοντέλου. Βασιζόμενοι στην μεθοδολογία που εισήχθη από τον Carter το 1992 [11], ανέπτυξαν μία μεθοδολογία που επιτρέπει, συνδυάζοντας την

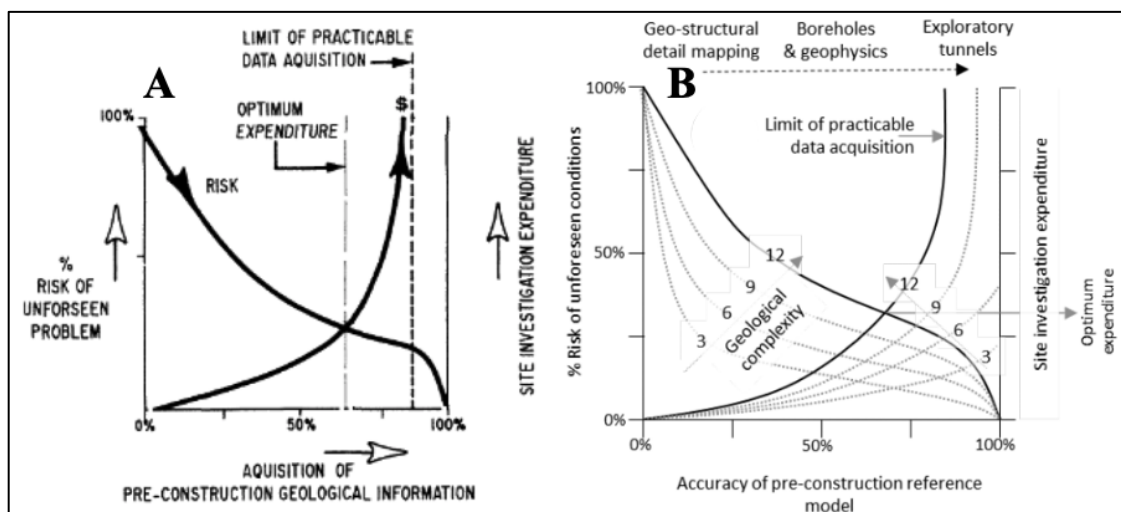
πολυπλοκότητα του γεωλογικού μοντέλου και την έκταση της έρευνας του υπεδάφους, να εξαχθεί ο βαθμός ακρίβειας του παραγόμενου γεωτεχνικού μοντέλου. Στο Σχήμα 2 φαίνεται ο πίνακας μέσω του οποίου προσδιορίζεται η πολυπλοκότητα του γεωλογικού μοντέλου, υπό την μορφή κλίμακας με τιμές από το 1 έως το 12 και στο Σχήμα 3 το διάγραμμα μέσω του οποίου προσδιορίζεται η ποσοστιαία ακρίβεια του γεωτεχνικού μοντέλου.

A. Simplified numerical representation of the geological complexity						
Geological complexity (simplified from 1 to 12)						
	Sedimentary		Magmatic		Metamorphic	
	Simple	Difficult	Simple	Difficult	Simple	Difficult
Unfolded	1,2	3	2	3		
Folded	3	4	4	5,6	5	7
Folded + Faulted	4	5	6,7	7,8	8,9	10,11
Folded + Faulted + Sheared		6,7		9	9	12

B. Uncertainty of the different models after a standard site investigation expenditure						
Uncertainty (after standard site investigation expenditure)						
	Sedimentary		Magmatic		Metamorphic	
	Simple	Difficult	Simple	Difficult	Simple	Difficult
Unfolded	VL	LM	L	LM		
Folded	LM	LM	LM	M	M	H
Folded + Faulted	LM	M	MH	H	H	VH
Folded + Faulted + Shared		MH		HVH	HVH	EH

VL = very low; L = low; LM = low-medium; M = medium; MH = medium high; H = high; HVH = high- very high; VH = very high; EH = extremely high

Σχήμα 2. Πίνακας προσδιορισμού πολυπλοκότητας γεωλογικού μοντέλου και αβεβαιότητας μετά από ορισμένη έκταση έρευνας πεδίου [10].

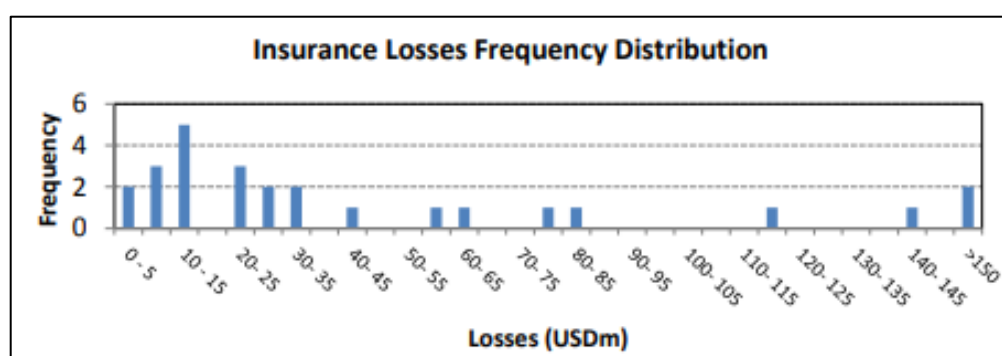


Σχήμα 3. Διάγραμμα προσδιορισμού ακρίβειας γεωτεχνικού μοντέλου [2] [3].

2.2 Αστοχίες – Απώλειες

Τις τελευταίες δεκαετίες, με την επέκταση και άνθιση του κατασκευαστικού τομέα σηράγγων, μέσω της υλοποίησης σημαντικών έργων σε όλο τον κόσμο, αντιμετωπίστηκαν και νέες προκλήσεις. Ο αυξημένος ανταγωνισμός της αγοράς, η κλίμακα μεγέθους των έργων και η πολυπλοκότητά τους, αλλά και η λήψη αποφάσεων με πολιτικά κριτήρια, οδήγησε στην εκπόνηση ιδιαίτερα πειστικών χρονοδιαγραμμάτων και προϋπολογισμών. Ως εκ τούτου, η πίεση για την έγκαιρη ολοκλήρωση των έργων προς αποφυγή επιβολής ρητρών σε συνδυασμό με την προσπάθεια μείωσης του κόστους μέσω εφαρμογής καινοτόμων κατασκευαστικών τεχνικών, οδήγησε στην αύξηση της πιθανότητας πρόκλησης περιστατικών αστοχίας. Αυτό με τη σειρά του οδήγησε στην άσκηση επιπλέον πίεσης στους κατασκευαστές αλλά και στην ασφαλιστική αγορά που βίωσε σημαντικές οικονομικές απώλειες [5].

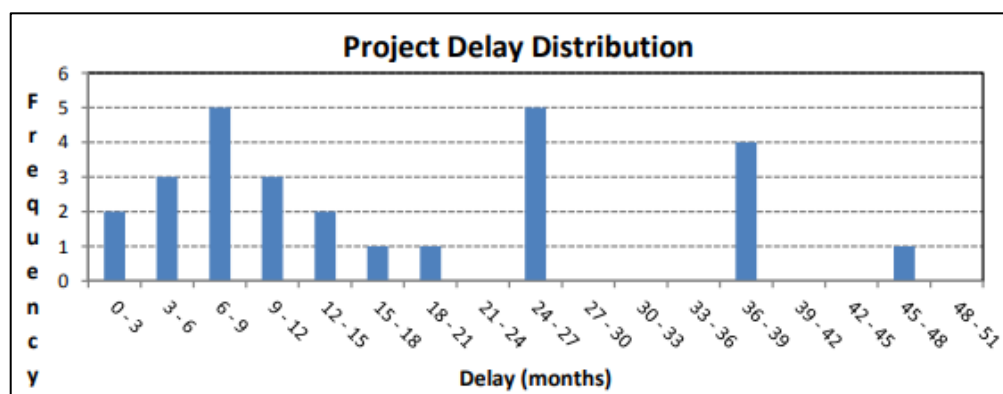
Στο Σχήμα 4 απεικονίζεται η κατανομή συχνότητας των οικονομικών απωλειών που καταγράφηκαν στην ασφαλιστική αγορά, σύμφωνα με τους Konstantis et al., για ένα περιορισμένο δείγμα έργων καθότι τα οικονομικά στοιχεία θεωρούνται ευαίσθητα δεδομένα και δεν διατίθενται στο ευρύ κοινό. Οι οικονομικές απώλειες φαίνεται να καλύπτουν ένα πολύ μεγάλο εύρος τιμών και σε ορισμένες περιπτώσεις να ξεπερνούν τα 150 εκ. δολάρια. Πάραυτα, παρατηρείται μία σχεδόν κανονική κατανομή μέχρι την τιμή των 50 εκ. δολαρίων, με μία μέση τιμή 10 - 15 εκ. δολαρίων και ένα διευρυμένο ανώτατο όριο που ξεπερνά τα 50 εκ. και κορυφώνεται στην περιοχή των 100+ εκ. δολαρίων [5].



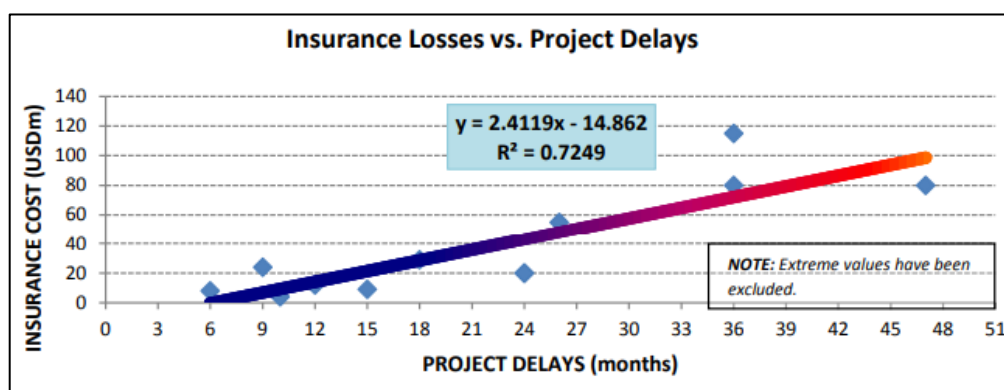
Σχήμα 4. Κατανομή οικονομικών απωλειών ασφαλιστικής αγοράς για την κάλυψη έργων κατασκευής σηράγγων [5].

Βασιζόμενοι στα ίδια περιορισμένα δεδομένα που αφορούν και στις αντίστοιχες χρονικές καθυστερήσεις που καταγράφηκαν μέχρι την ολοκλήρωση των έργων, οι

Konstantis et al. διαπίστωσαν, όπως είναι εμφανές και από την κατανομή που απεικονίζεται στο Σχήμα 5, ότι η έκταση των χρονικών καθυστερήσεων ποικίλει και ξεκινάει από ιδιαίτερα χαμηλές τιμές (π.χ. 1 μήνας) έως εξαιρετικά υψηλές (π.χ. 4 χρόνια). Πάραυτα, οι χρονικές καθυστερήσεις παρουσιάζουν όμοια συμπεριφορά με τις οικονομικές απώλειες, παρουσιάζοντας μία σχετικά κανονική κατανομή μέχρι σχεδόν την τιμή του 1 έτους. Καθυστερήσεις της τάξεως των 3 ετών φαίνεται να είναι αρκετά συχνές, ενώ μεγαλύτερες καθυστερήσεις θεωρούνται εξαιρετικά σπάνιες. Στο Σχήμα 6 απεικονίζεται μια προσπάθεια γραμμικής συσχέτισης των οικονομικών απωλειών και των χρονικών καθυστερήσεων, μη λαμβάνοντας υπόψη τις ακραίες περιπτώσεις [5].



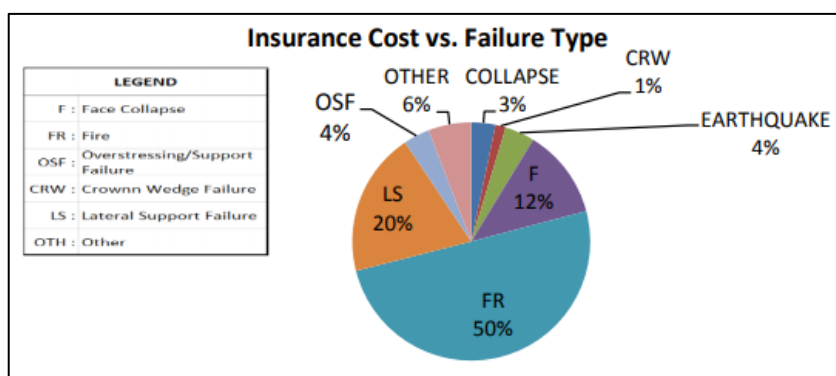
Σχήμα 5. Κατανομή χρονικών καθυστερήσεων έργων κατασκευής σηράγγων [5].



Σχήμα 6. Συσχέτιση οικονομικών απωλειών και χρονικής καθυστέρησης έργων [5].

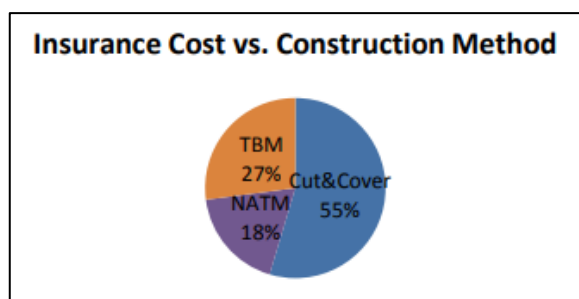
Στο Σχήμα 7 απεικονίζονται οι οικονομικές απώλειες σε συσχέτιση με τους τύπους αστοχίας που καταγράφηκαν για το ίδιο δείγμα έργων. Όπως φαίνεται, σε ποσοστό 50% οι απώλειες οφείλονται σε εκδήλωση πυρκαγιάς στις σήραγγες, ιδίως κατά τη φάση

λειτουργίας τους. Χαρακτηριστικά παραδείγματα είναι η πυρκαγιά στο Eurotunnel (τόσο το 1996 όσο και το 2008), στη σήραγγα Montblanc στη Γαλλία το 1999 και στο Tauertunnel στην Αυστρία το 1999. Το ένα πέμπτο περίπου των συνολικών απωλειών οφείλονται σε αστοχίες λόγω κατάρρευσης των τοίχων αντιστήριξης, με πιο χαρακτηριστικές περιπτώσεις αυτές του αυτοκινητοδρόμου MRT / Niccol το 2004 και του μετρό της Κολωνίας το 2009. Όσον αφορά στην αστοχία σηράγγων κατά την φάση κατασκευής, η αστάθεια μετώπου φαίνεται ως ο τύπος αστοχίας που οδηγεί στις μεγαλύτερες απώλειες [5].



Σχήμα 7. Οικονομικές απώλειες ανά τύπο αστοχίας [5].

Στο Σχήμα 8 απεικονίζεται η συσχέτιση των οικονομικών απωλειών με την μέθοδο κατασκευής των σηράγγων του δείγματος. Το υψηλότερο ποσοστό απωλειών συσχετίζεται με τη μέθοδο Cut & Cover, ενώ ακολουθεί αυτό που συσχετίζεται με τη διάνοιξη με τη χρήση TBM και τελευταίο αυτό που συσχετίζεται με την μέθοδο NATM [5].



Σχήμα 8. Οικονομικές απώλειες ανά μέθοδο κατασκευής [5].

3 Μηχανική Μάθηση

3.1 Ορισμός Μηχανικής Μάθησης

Η διαδικασία μέσω της οποίας ο άνθρωπος κατανοεί το περιβάλλον του, συνίσταται από την παρατήρηση και τη δημιουργία μιας απλοποιημένης (αφαιρετικής) εκδοχής του, η οποία ονομάζεται *μοντέλο (model)*. Η καθαυτή δημιουργία του μοντέλου, ονομάζεται *επαγωγική μάθηση (inductive learning)* και η διαδικασία στην οποία εμπεριέχεται, ονομάζεται *επαγωγή (induction)*. Ο τρόπος με τον οποίο ο άνθρωπος οργανώνει τις εμπειρίες του, είναι μέσω της δημιουργίας δομών συσχέτισης αυτών, που ονομάζονται *πρότυπα (patterns)*. Η αντίστοιχη διαδικασία μέσω της οποίας τα υπολογιστικά συστήματα δημιουργούν μοντέλα ή πρότυπα χρησιμοποιώντας ένα σύνολο δεδομένων, ονομάζεται *μηχανική μάθηση (machine learning)* [12].

3.2 Είδη Μηχανικής Μάθησης

Αναλόγως τη φύση του προβλήματος, χρησιμοποιούνται διάφορες τεχνικές μηχανικής μάθησης, οι οποίες ανήκουν στις δύο παρακάτω κατηγορίες [12]:

α. *Επιβλεπόμενη μάθηση (supervised learning)* ή *μάθηση μέσω παραδειγμάτων (learning from examples)*.

β. *Μη επιβλεπόμενη μάθηση (unsupervised learning)* ή *μάθηση μέσω παρατήρησης (learning from observation)*.

Στην επιβλεπόμενη μάθηση το υπολογιστικό σύστημα “μαθαίνει” μια έννοια ή συνάρτηση, η οποία αποτελεί την περιγραφή του μοντέλου, χρησιμοποιώντας ένα σύνολο δεδομένων. Ο όρος επιβλεπόμενη προκύπτει καθότι θεωρείται ότι υπάρχει κάποιος ο οποίος “επιβλέπει” τη διαδικασία, παρέχοντας μια δεδομένη τιμή εξόδου για τη συνάρτηση, για τα υπό εξέταση δεδομένα.

Αντιθέτως, στη μη επιβλεπόμενη μάθηση, το υπολογιστικό σύστημα ανακαλύπτει μόνο του τις συσχετίσεις ή τις ομάδες που προκύπτουν από τα παρεχόμενα δεδομένα, δημιουργώντας *πρότυπα (patterns)*, χωρίς να γνωρίζει αν υπάρχουν, ποια είναι και πόσα είναι. Τα μοντέλα χαρακτηρίζονται ως *μοντέλα πρόβλεψης (predictive models)*, καθότι

προβλέπουν την τιμή κάποιας άγνωστης μεταβλητής και περιγράφουν συνολικά τα δεδομένα. Πέρα από τη δυνατότητα πρόβλεψης, υπάρχει επιπλέον και δυνατότητα παροχής ποιοτικών πληροφοριών για τα δεδομένα. Τα πρότυπα, αντιθέτως, περιγράφουν ένα τμήμα των δεδομένων και χαρακτηρίζονται ως *πρότυπα πληροφόρησης (informative patterns)*, καθότι περιγράφουν τις συσχετίσεις των δεδομένων.

3.3 Μη Επιβλεπόμενη Μάθηση

Στη μη επιβλεπόμενη μάθηση το σύστημα στοχεύει στην ανακάλυψη συσχετίσεων και ομάδων από τα δεδομένα, βασιζόμενο στις ιδιότητες τους. Ως εκ τούτου, κάθε πρότυπο (περιγραφή) που προκύπτει αφορά σε ένα τμήμα των δεδομένων. Οι *ομάδες - συστάδες (clusters)* και οι *κανόνες συσχέτισης (association rules)*, αποτελούν παραδείγματα προτύπων τα οποία προκύπτουν από τη διαδικασία της *συσταδοποίησης (clustering)* ή *ομαδοποίησης* [12].

3.3.1 Συσταδοποίηση

Η συσταδοποίηση (clustering) θεωρείται μία από τις χρησιμότερες διεργασίες εξόρυξης γνώσης από βάσεις δεδομένων, οι οποία οδηγεί στην ανακάλυψη συστάδων και στον προσδιορισμό προτύπων (patterns) ή κατανομών στα εξεταζόμενα δεδομένα. Η συσταδοποίηση οδηγεί στην *τμηματοποίηση (partitioning)* ενός συνόλου δεδομένων σε ομάδες ή συστάδες, ώστε τα στοιχεία μίας συστάδας να παρουσιάζουν μεγαλύτερη ομοιότητα, έναντι της ομοιότητας που παρουσιάζουν με τα στοιχεία οποιασδήποτε άλλης συστάδας [13] [14] [15]. Χαρακτηριστικό παράδειγμα, είναι η διαδικασία συσταδοποίησης ή ομαδοποίησης των πελατών ενός καταστήματος λιανικής πώλησης. Σε αυτή την περίπτωση οι πελάτες με όμοια αγοραστικά ενδιαφέροντα ή αγοραστικές προτιμήσεις, δηλαδή με ίδια αγοραστικά πρότυπα, ανήκουν στην ίδια ομάδα – συστάδα. Βασικό μέλημα λοιπόν της συσταδοποίησης, είναι η αποκάλυψη της οργάνωσης προτύπων στις συστάδες, η οποία μας επιτρέπει να ανακαλύψουμε τις διαφορές και τις ομοιότητες αυτών και να εξάγουμε χρήσιμα συμπεράσματα για τα δεδομένα μας. Αναλόγως του πεδίου εφαρμογής της, η συσταδοποίηση απαντάται με διαφορετικά ονόματα όπως *μη εποπτευόμενη μάθηση (unsupervised learning)* στην αναγνώριση προτύπων, *αριθμητική ταξονομία (numerical*

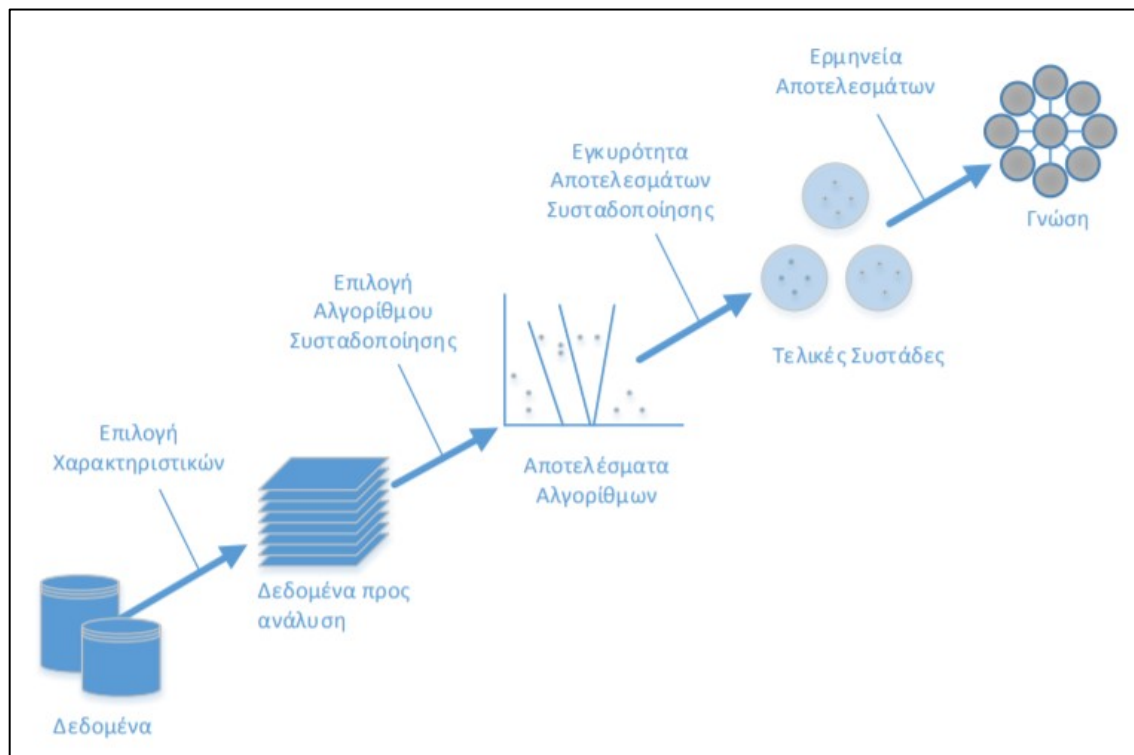
taxonomy) στην βιολογία - οικολογία, *τυπολογία (typology)* στις κοινωνικές επιστήμες και *τμηματοποίηση (partitioning)* στη θεωρία των γραφών [13] [16].

Αναλόγως του κριτηρίου που εφαρμόζεται για τη συσταδοποίηση ενός συνόλου δεδομένου, δύναται να προκύψει διαφορετική τμηματοποίηση. Συνεπώς, πριν από τη εφαρμογή της διαδικασίας συσταδοποίησης απαιτείται η προεπεξεργασία των δεδομένων.

Στο Σχήμα 9 φαίνονται τα βασικά βήματα ανάπτυξης της διαδικασίας συσταδοποίησης, τα οποία συνοψίζονται στα παρακάτω [13]:

α. *Επιλογή χαρακτηριστικών γνωρισμάτων.* Αναλόγως την εργασία που μας ενδιαφέρει να εκτελέσουμε, επιλέγουμε τα κατάλληλα *χαρακτηριστικά – γνωρίσματα (features – attributes)* ώστε να επιτυγχάνεται η μέγιστη δυνατή κωδικοποίηση πληροφορίας. Επομένως, πριν από τη χρήση των δεδομένων για την εφαρμογή της διαδικασίας συσταδοποίησης, μπορεί να απαιτείται η προεπεξεργασία τους.

β. *Επιλογή αλγορίθμου συσταδοποίησης.* Αυτό το βήμα αφορά στην επιλογή του κατάλληλου αλγορίθμου που θα οδηγήσει σε ένα καλό σχήμα συσταδοποίησης (*clustering scheme*). Το *κριτήριο συσταδοποίησης* και το *μέτρο γειτνίασης* είναι στοιχεία του αλγορίθμου συσταδοποίησης που χαρακτηρίζουν τη δυνατότητα καθορισμού ενός σχήματος συσταδοποίησης κατάλληλου για το σύνολο των δεδομένων.



Σχήμα 9. Βήματα της διαδικασίας συσταδοποίησης [13].

i. *Κριτήριο συσταδοποίησης*. Η έκφραση του κριτηρίου μπορεί να πραγματοποιηθεί μέσω μίας συνάρτησης κόστους ή μέσω κάποιου άλλου τύπου κανόνα. Πρέπει να τονιστεί ότι είναι σημαντικό να λαμβάνεται υπόψιν κατά την επιλογή του κριτηρίου, ο τύπος των συστάδων που αναμένεται να εμφανιστούν και να επιλέγεται το κατάλληλο κριτήριο που δύναται να οδηγήσει σε τμηματοποίηση που να ταιριάζει στο σύνολο των δεδομένων.

ii. *Μέτρο γειτνίασης*. Το μέτρο αυτό προσδιορίζει πόσο όμοια είναι δύο αντικείμενα, δηλαδή δύο εγγραφές της βάσης, με τα αντίστοιχα διανύσματα γνωρισμάτων – χαρακτηριστικών που προκύπτουν. Σε πολλές περιπτώσεις, πρέπει να εξασφαλίζεται ότι η βαρύτητα συμβολής του κάθε γνωρίσματος – χαρακτηριστικού στο μέτρο γειτνίασης είναι ισότιμη και κανένα γνώρισμα δεν παρουσιάζει μεγαλύτερη βαρύτητα.

γ. *Επικύρωση αποτελεσμάτων*. Εφόσον τα αποτελέσματα της διαδικασίας συσταδοποίησης δεν είναι εκ των προτέρων γνωστά, απαιτείται η αξιολόγηση των αποτελεσμάτων και η εξακρίβωση της ακρίβειας τους, η οποία πραγματοποιείται χρησιμοποιώντας κατάλληλα κριτήρια – τεχνικές και οδηγεί στην τελική τμηματοποίηση του συνόλου των δεδομένων [17].

δ. *Ερμηνεία των αποτελεσμάτων*. Η εμπειρία στην περιοχή εφαρμογής της διαδικασίας συσταδοποίησης είναι απαραίτητη προκειμένου να υπάρχει δυνατότητα ερμηνείας των αποτελεσμάτων, τα οποία πολλές φορές απαιτείται να συνδυαστούν με πειραματικά στοιχεία και αποτελέσματα προηγούμενων αναλύσεων, ώστε να προκύψουν τα σωστά συμπεράσματα.

3.3.2 Κατηγοριοποίηση Αλγορίθμων Συσταδοποίησης

Στη βιβλιογραφία έχει προταθεί ένα πλήθος αλγορίθμων οι οποίοι διαφοροποιούνται ανάλογα με τη μέθοδο που υιοθετείται για τον ορισμό των συστάδων και οι οποίοι μπορούν να ταξινομηθούν ως κάτωθι [13]:

α. *Διαιρετική συσταδοποίηση (partitional clustering)*. Γενικό κριτήριο σε αυτήν την κατηγορία είναι η μεγιστοποίηση κάποιου μέτρου ανομοιότητας μεταξύ των συστάδων και η ελαχιστοποίηση κάποιου μέτρου ανομοιότητας μεταξύ των στοιχείων της ίδιας συστάδας.

β. *Ασαφής συσταδοποίηση (fuzzy clustering)*. Σε αυτήν την κατηγορία θεωρείται ότι ένα στοιχείο δύναται να ταξινομηθεί σε παραπάνω από μία συστάδες και για την ομαδοποίηση των δεδομένων χρησιμοποιούνται τεχνικές ασαφούς λογικής.

γ. *Μη ασαφής συσταδοποίηση (crisp clustering)*. Θεωρεί ότι ένα στοιχείο μπορεί να ανήκει μόνο σε μία συστάδα και για τον διαχωρισμό των συστάδων θεωρούνται μη επικαλυπτόμενα χωρίσματα. Οι περισσότεροι αλγόριθμοι συσταδοποίησης οδηγούν σε σαφείς συστάδες και επομένως μπορούν να ανήκουν σε αυτήν την κατηγορία.

δ. *Συσταδοποίηση βασισμένη στα δίκτυα Kohonen (Kohonen net clustering)*. Αυτή η κατηγορία βασίζεται σε έννοιες των νευρωνικών δικτύων. Η συσταδοποίηση πραγματοποιείται μέσω χρήσης αλγορίθμων, οι οποίοι αλλάζουν τα βάρη μεταξύ των συνδέσεων του δικτύου, ώστε οι κόμβοι εξόδου να τείνουν να σχηματίσουν συστάδες.

ε. *Ιεραρχική Συσταδοποίηση (hierarchical Clustering)*. Σε αυτήν την κατηγορία οι αλγόριθμοι συνδέουν διαδοχικά μικρότερες συστάδες σε μεγαλύτερες ή διασπούν μεγαλύτερες συστάδες σε μικρότερες και διαφέρουν μεταξύ τους ως προς τον κανόνα – κριτήριο, με το οποίο αποφασίζεται ποιες συστάδες θα συγχωνευτούν ή θα διασπαστούν.

3.3.3 Αλγόριθμος K-Means

Ο αλγόριθμος K-Means αποτελεί έναν από τους πιο συχνά χρησιμοποιούμενους αλγορίθμους συσταδοποίησης [18]. Ανήκει στην κατηγορία αλγορίθμων διαιρετικής συσταδοποίησης και βασίζεται στην αποσύνθεση του συνόλου δεδομένων σε ασυσχέτιστες συστάδες. Ο αλγόριθμος προσπαθεί να ελαχιστοποιήσει την συνάρτηση μέσης τετραγωνικής ευκλείδειας απόστασης των σημείων που ορίζονται από τα στοιχεία – εγγραφές της βάσης δεδομένων, από τα πλησιέστερα κέντρα (*κεντροειδή*) των συστάδων και δίνεται από την **Εξίσωση 1**:

$$E = \sum_{i=1}^k \sum_{x \in C_i} d(X, M_i)$$

Εξίσωση 1

Στην Εξίσωση 1, M_i είναι το κεντροειδές της συστάδας C_i και $d(X, M_i)$ η τετραγωνική ευκλείδεια απόσταση μεταξύ του κεντροειδούς M_i και ενός σημείου X (μία

εγγραφή της βάσης δεδομένων) που περιγράφεται από τις αριθμητικές παραμέτρους A_1, A_2, \dots, A_p (οι τιμές των χαρακτηριστικών – γνωρισμάτων), η οποία δίνεται από την **Εξίσωση 2**:

$$d(X, M) = \sum_{j=1}^p (X_j - M_j)^2$$

Εξίσωση 2

Επομένως, ο αλγόριθμος προσπαθεί να ελαχιστοποιήσει την συνάρτηση E , δηλαδή το άθροισμα των τετραγωνικών ευκλείδειων αποστάσεων των σημείων X , από τα κεντροειδή των συστάδων. Ο αριθμός k είναι ο αριθμός των συστάδων στις οποίες επιθυμούμε να διαχωρίσουμε το σύνολο των δεδομένων και ως εκ τούτου και ο αριθμός των κεντροειδών. Ο αλγόριθμος ξεκινά με έναν αρχικό ορισμό των κεντροειδών (*αρχικοποίηση*) και στη συνέχεια υπολογίζοντας την τετραγωνική ευκλείδεια απόσταση του κάθε στοιχείου του συνόλου δεδομένων από τα κεντροειδή, αναθέτει κάθε στοιχείο στη συστάδα της οποίας το κεντροειδές βρίσκεται πιο κοντά. Στη συνέχεια υπολογίζονται εκ νέου τα κεντροειδή, με συντεταγμένες το μέσο όρο των συντεταγμένων των σημείων που ανήκουν στη συστάδα, υπολογίζονται εκ νέου οι τετραγωνικές ευκλείδειες αποστάσεις των σημείων από τα κεντροειδή, ώστε να ανατεθεί εκ νέου κάθε στοιχείο στη συστάδα, με βάση το κριτήριο ελάχιστης τετραγωνικής ευκλείδειας απόστασης από το κεντροειδές της. Η διαδικασία επαναλαμβάνεται μέχρις ότου να πάψουν να μεταβάλλονται οι συστάδες και η τιμή της συνάρτησης E να παραμένει (σχετικά) σταθερή [13].

3.3.4 Αλγόριθμος K-Prototypes

Ο αλγόριθμος K-Prototypes σχεδιάστηκε με σκοπό τη συσταδοποίηση συνόλων δεδομένων που εμπεριέχουν αριθμητικά και λεκτικά δεδομένα. Ο αλγόριθμος αυτός βασίζεται στον K-Means, είναι επίσης αλγόριθμος διαιρετικής συσταδοποίησης και θεωρείται πιο χρήσιμος, διότι οι βάσεις δεδομένων που χρησιμοποιούνται κατά κύριο λόγο για πρακτικές εφαρμογές, αποτελούνται από εγγραφές μικτού τύπου. Το μέτρο ανομοιότητας μεταξύ δύο εγγραφών μικτού τύπου X και Y , που περιγράφονται από τις αριθμητικές παραμέτρους A_1, A_2, \dots, A_p και τις περιγραφικές παραμέτρους A_{p+1}, \dots, A_m , δίνεται από την **Εξίσωση 3** [19]:

$$d_2(X, Y) = \sum_{j=1}^p (X_j - Y_j)^2 + \gamma \sum_{j=p+1}^m \delta(X_j, Y_j)$$

Εξίσωση 3

Στην Εξίσωση 3 ο πρώτος όρος είναι η τετραγωνική ευκλείδεια απόσταση των εγγραφών X , Y και ο δεύτερος όρος είναι το γινόμενο του μέτρου ανομοιότητας (*dissimilarity measure*) των εγγραφών X , Y με έναν συντελεστή γ που χρησιμοποιείται για την εξισορρόπηση της βαρύτητας των δύο όρων.

Η διαδικασία της συσταδοποίησης του αλγορίθμου K-Prototypes είναι όμοια με του K-Means, με τη διαφορά ότι πρέπει να δίνεται προσοχή στην κατάλληλη επιλογή της τιμής του συντελεστή γ , ώστε να αποφεύγεται η εύνοια κάποιου τύπου γνωρισμάτων [13].

3.3.4.1 Μέτρο Ανομοιότητας

Έστω X , Y δύο εγγραφές της βάσης δεδομένων με m περιγραφικά – λεκτικά γνωρίσματα. Με βάση την ανομοιότητα που παρουσιάζουν οι δύο εγγραφές στις τιμές των περιγραφικών γνωρισμάτων, ορίζεται η συνολική ανομοιότητα των εγγραφών. Όσο μεγαλύτερος είναι ο αριθμός των αταίριαστων τιμών για τα αντίστοιχα γνωρίσματα των δύο εγγραφών, τόσο ανόμοιες θεωρούνται οι δύο εγγραφές. Τυπικά, το μέτρο ανομοιότητας εκφράζεται από την Εξίσωση 4:

$$d(X, Y) = \sum_{j=1}^m \delta(X_j, Y_j)$$

Εξίσωση 4

, όπου:

$$\delta(X_j, Y_j) = \begin{cases} 0 & (X_j = Y_j) \\ 1 & (X_j \neq Y_j) \end{cases}$$

Εξίσωση 5

3.4 Εφαρμογές Μηχανικής Μάθησης για την Εκτίμηση Κινδύνου στην Κατασκευή Σηράγγων

Τα τελευταία χρόνια η μηχανική μάθηση βοήθησε στη επίλυση διαφόρων θεμάτων μηχανικής, από την ανίχνευση ελαττωμάτων στο οπλισμένο σκυρόδεμα και την παρακολούθηση φυσικών καταστροφών, έως την επεξεργασία σε πραγματικό χρόνο δεδομένων για την πλοήγηση μη επανδρωμένων αεροσκαφών [20]. Το ερώτημα όμως που ανακύπτει ως προς το αντικείμενο της παρούσας εργασίας, είναι αν η μηχανική μάθηση δύναται να συνεισφέρει στην εκτίμηση και την διαχείριση κινδύνου κατά την κατασκευή σηράγγων. Και ως εκτίμηση κινδύνου σε αυτή την περίπτωση, θεωρείται η συνολική διαδικασία αναγνώρισης του κινδύνου, ανάλυσής, αξιολόγησης και πιθανόν διαχείρισης [20]. Η απάντηση σε αυτό το ερώτημα, προκύπτει μέσω αναζήτησης σχετικών εφαρμογών στη διεθνή βιβλιογραφία. Παρακάτω αναφέρονται ορισμένα παραδείγματα τέτοιων εφαρμογών.

Οι Benardos και Kaliampakos το 2004, χρησιμοποιώντας δεδομένα από την κατασκευή του μετρό της Αθήνας, μοντελοποίησαν την απόδοση TBM, μέσω της ανάπτυξης ενός τεχνητού νευρωνικού δικτύου (artificial neural network, ANN). Το μοντέλο πέρα από την ικανότητά του να εξάγει εκτιμήσεις για την προχώρηση της διάνοιξης, παρέχει τη δυνατότητα εντοπισμού συνθηκών εδάφους που ενδέχεται να διαταράξουν τον κύκλο λειτουργίας και την απόδοση του TBM [21]. Η δυνατότητα προσδιορισμού περιοχών που είναι επιρρεπείς σε κίνδυνο είναι πολύτιμη γνώση η οποία, στη φάση σχεδιασμού του έργου, καθιστά δυνατή την επιλογή της κατάλληλης όδευσης της σήραγγας, των χαρακτηριστικών του TBM ή/και την κατάλληλη τεχνική βελτίωσης του εδάφους [22]. Επιπλέον είναι εφικτός ο εκ των προτέρων σχεδιασμός και η ανάπτυξη ενός συστήματος διαχείρισης κινδύνου, το οποίο διασφαλίζει την οικονομική και την προγραμματική αξιοπιστία του έργου [21].

Οι Sousa και Einstein το 2012, ανέπτυξαν μία μεθοδολογία για τη συστηματική αξιολόγηση και διαχείριση των κινδύνων που σχετίζονται με την κατασκευή σηράγγων. Η μεθοδολογία συνίσταται στο συνδυασμό δύο μοντέλων μηχανική μάθησης, ενός μοντέλου πρόβλεψης της γεωλογίας του περιβάλλοντος κατασκευή της σήραγγας, με ένα μοντέλο επιλογής κατασκευαστικής στρατηγικής, που επιτρέπει σε κάποιον να επιλέξει ανάμεσα σε διαφορετικές στρατηγικές εκείνη που οδηγεί στον ελάχιστο κίνδυνο. Το μοντέλο πρόβλεψης γεωλογίας χρησιμοποιεί δεδομένα απόδοσης TBM, ενώ και τα δύο μοντέλα χρησιμοποιούν Bayesian δίκτυα (Bayesian Networks) για την διαδικασία ανάλυσης των δεδομένων. Το

συνδυασμένο μοντέλο εφαρμόστηκε κατά την κατασκευή του Porto Metro, στην Πορτογαλία. Τα αποτελέσματα του μοντέλου γεωλογικής πρόβλεψης ήταν σύμφωνα με την γεωλογία που παρατηρήθηκε στην πραγματικότητα και τα αποτελέσματα του μοντέλου επιλογής αποφάσεων κατασκευής ήταν σύμφωνα με τις μεθόδους και την στρατηγική που ακολουθήθηκε στην πραγματικότητα. Ως πολύ σημαντική αξιολογήθηκε η ικανότητα του μοντέλου να προβλέπει αλλαγές στη γεωλογία και να υποδεικνύει την απαίτηση αλλαγής της κατασκευαστικής στρατηγικής. Το εν λόγω μοντέλο εκτίμησης κινδύνου είναι ένα ισχυρό εργαλείο με το οποίο οι σχεδιαστές και οι μηχανικοί μπορούν να αξιολογήσουν και να μετριάσουν τους κινδύνους που σχετίζονται με την κατασκευή σιράγγων. [23]

Οι Ding και Zhou το 2013 ανέπτυξαν ένα σύστημα, που με τη χρήση τεχνητών νευρωνικών δικτύων (artificial neural networks, ANN), αξιοποιεί δεδομένα που καταγράφονται σε πραγματικό χρόνο κατά την κατασκευή αστικών γραμμών μετρό, πραγματοποιεί εκτίμηση και αξιολόγηση κινδύνου και εξάγει έγκαιρα και αυτοματοποιημένα προειδοποιήσεις για την ασφάλεια του έργου έναντι εκδήλωσης αστοχιών. Το σύστημα χρησιμοποιήθηκε επιτυχώς κατά την κατασκευή διαφόρων γραμμών μετρό (Wuhan, Shenyang, Zhengzhou, Kunming), όπου παρατηρήθηκε σημαντική βελτίωση στην διαχείριση του κινδύνου κατά την εξέλιξη των έργων. [24]

Οι Hyun et al. το 2015 πρότειναν μία μέθοδο εκτίμησης και αξιολόγησης κινδύνου που αφορά στη λειτουργία του TBM κατά τη διάνοιξη σήραγγας. Με βάση στοιχεία από έργα που πραγματοποιήθηκαν με τη χρήση TBM, κατηγοριοποίησαν τα πιθανά προβλήματα σε τέσσερις κατηγορίες (δυσλειτουργία κοπτικών, μπλοκάρισμα ή εγκλωβισμό του TBM, προβλήματα στην μεταφορά των υλικών εκσκαφής και προβλήματα που αφορούν στη τοποθέτηση της επένδυσης) και ομαδοποίησαν τους παράγοντες κινδύνου (γεωλογικούς, σχεδιαστικούς, κατασκευαστικούς/διαχειριστικούς). Χρησιμοποιώντας μοντέλα μηχανικής μάθησης που ενσωματώνουν δέντρα σφαλμάτων (fault tree analysis, FTA) και λαμβάνοντας υπόψη την πιθανότητα εκδήλωσης των συμβάντων και τις επιπτώσεις τους, εξήγαγαν αποτελέσματα εκτίμησης κινδύνου. Τα αποτελέσματα της προτεινόμενης μεθόδου κρίθηκαν ως αξιόπιστα, κατόπιν σύγκρισης τους με πραγματικά στοιχεία έργων και η μέθοδος χαρακτηρίστηκε ως κατάλληλη για εφαρμογή στην πράξη. [25]

Οι Mikaeli et al. το 2019 ανέπτυξαν μοντέλα μηχανικής μάθησης για την αξιολόγηση και ταξινόμηση των γεωτεχνικών κινδύνων που αντιμετωπίζονται κατά τη διάνοιξη σιράγγων σε μεγάλα βάθη. Τα μοντέλα ενσωματώνοντας τεχνικές συσταδοποίησης

(clustering techniques) και αξιοποιώντας δεδομένα από την κατασκευή της σήραγγας Emamzade Hashem στο Ιράν, εξάγουν αποτελέσματα βέλτιστης διαχείρισης κινδύνου. Τα δεδομένα αφορούσαν στην ύπαρξη περιοχών αστάθειας περιμετρικά της σήραγγας, στην εισροή υδάτων και στην ύπαρξη συνθηκών σύνθλιψης και χρησιμοποιήθηκαν ως κριτήρια ανάλυσης για την εκτίμηση - αξιολόγηση του κινδύνου. Από τον έλεγχο αποτελεσματικότητας των μοντέλων με βάση τα πραγματικά στοιχεία, κρίθηκε ότι τα μοντέλα παρουσίασαν σημαντική ικανότητα αξιολόγησης των κινδύνων. [26]

4 Δεδομένα προς Ανάλυση

4.1 Περιγραφή Βάσης Δεδομένων

Τα δεδομένα τα οποία χρησιμοποιήθηκαν για την εφαρμογή των μοντέλων, προέρχονται από τη βάση δεδομένων η οποία δημιουργήθηκε από τους Konstantis, et al. το 2016 [5] και στη συνέχεια συμπληρώθηκε και επεκτάθηκε από τον Dallavalle το 2019 [27].

Η αρχική βάση δεδομένων περιέχει στοιχεία για αξιοσημείωτα περιστατικά αστοχίας σηράγγων που έχουν καταγραφεί παγκοσμίως από το έτος 1941 έως το 2019, τα οποία συντελέστηκαν είτε κατά την χρονική περίοδο κατασκευής, είτε κατά την χρονική περίοδο λειτουργίας των σηράγγων. Η υπόψη βάση δεδομένων συμπληρώθηκε μερικώς και επεκτάθηκε κατά το έτος 2019, λαμβάνοντας υπόψη όμως μόνο περιστατικά αστοχίας που καταγράφηκαν μετά το έτος 1980, καθότι θεωρήθηκε ότι αυτά τα περιστατικά αφορούν σε αστοχίες σχετιζόμενες με τις σύγχρονες μεθόδους διάνοιξης σηράγγων.

Επομένως, η βάση δεδομένων η οποία τελικώς παρασχέθηκε για την εκπόνηση της παρούσας εργασίας, αποτελείται από τριακόσιες εβδομήντα επτά (377) εγγραφές, μερικώς συμπληρωμένες, οι οποίες αφορούν σε αξιοσημείωτες αστοχίες σηράγγων που καταγράφηκαν παγκοσμίως μετά το έτος 1980. Τα στοιχεία - χαρακτηριστικά τα οποία συμπεριλαμβάνονται και τα αντίστοιχα ποσοστά πλήρωσής τους φαίνονται στον Πίνακα 1.

<i>Στοιχείο – Χαρακτηριστικό</i>	<i>Ποσοστό Πλήρωσης</i>
Year (Έτος)	75 %
Country (Χώρα)	100 %
Project (Έργο)	100 %
Tunnel Use (Χρήση)	79 %
Failure Type (Τύπος Αστοχίας)	97 %
Construction / Operation (Κατασκευή / Λειτουργία)	100 %
Cutting Technique (Μέθοδος Διάνοιξης)	62 %
Earthquake Seismicity (Σεισμικότητα)	0 %
Incident – Reports and Causes (Περιστατικό – Αναφορές και Αίτια)	61 %
Length (Μήκος) (m)	66 %
Diameter (Διάμετρος) (m)	38 %
Overburden (Υπερκείμενα) (m)	37 %
Rock / Soil (Βράχος / Έδαφος)	60 %
Geological Information (Γεωλογική Πληροφορία)	27 %
Stress Conditions (Έντατικό Πεδίο)	0 %
Water Conditions (Υδρογεωλογικές Συνθήκες)	2 %
Portal Failure (Αστοχία Στομίου)	1 %
Third Party Impact (Επιπτώσεις σε Τρίτους)	7 %
Fatalities (Ανθρώπινες Απώλειες)	14 %
Losses (Ζημιές) (mil. \$)	8 %
Delay (Καθυστέρηση)	6 %
Source of Information (Πηγή)	94 %

Πίνακας 1. Στοιχεία - χαρακτηριστικά περιστατικών αστοχίας σιράγγων που συμπεριλαμβάνονται στη βάση δεδομένων και αντίστοιχα ποσοστά πλήρωσης.

4.2 Προεπεξεργασία Δεδομένων

Η προεπεξεργασία των δεδομένων αφορά στις εργασίες προετοιμασίας των δεδομένων που πρέπει να πραγματοποιηθούν προκειμένου να επιτευχθεί η καθαυτή εξόρυξη γνώσης. Η προεπεξεργασία απαιτείται καθότι στα αρχικά δεδομένα δύναται να παρατηρηθούν διαφόρων ειδών προβλήματα, όπως η ύπαρξη αλληλοσυγκρουόμενων πληροφοριών, ασυνεπειών στην κωδικοποίηση, χαμένων τιμών και θορύβου κ.α. Η διαδικασία αντιμετώπισης αυτών των προβλημάτων καλείται *καθαρισμός δεδομένων*. Η προεπεξεργασία όμως δεν περιορίζεται μόνο στον καθαρισμό. Αναλόγως της χρησιμοποιούμενης μεθόδου επεξεργασίας, συχνά επιβάλλεται ο μετασχηματισμός των δεδομένων. Η *κανονικοποίηση* και η *διακριτοποίηση* είναι δύο συνήθεις εργασίες μετασχηματισμού, οι οποίες αφορούν αντιστοίχως στην μετατροπή των αριθμητικών τιμών σε άλλες “καταλληλότερες” αριθμητικές τιμές και στην μετατροπή - μετασχηματισμό αριθμητικών τιμών σε ονομαστικές τιμές. Ομοίως, αναλόγως της χρησιμοποιούμενης μεθόδου, μπορεί να απαιτηθεί η μετατροπή των ονομαστικών τιμών σε αριθμητικές. Ένα επιπλέον θέμα που αφορά στην προεπεξεργασία, είναι η μείωση του όγκου των δεδομένων. Από το σύνολο των δεδομένων ενδέχεται ορισμένες μόνο μεταβλητές ή πεδία να είναι απαραίτητα για την εξόρυξη γνώσης, επομένως η επιλογή των σημαντικών χαρακτηριστικών είναι μία εργασία βαρύνουσας σημασίας [28].

Προκειμένου να καταστεί δυνατή η εφαρμογή των μοντέλων που αναπτύχθηκαν κατά την παρούσα εργασία, προηγήθηκε προεπεξεργασία των δεδομένων, αποτελούμενη από τις παρακάτω εργασίες:

- α. Επιλογή χαρακτηριστικών.
- β. Πλήρωση χαμένων τιμών – προσθήκη νέων εγγραφών.
- γ. Διαχείριση θορυβωδών δεδομένων.
- δ. Μετατροπή ονομαστικών τιμών σε αριθμητικές (μόνο στο μοντέλο K-Means - TFM1).
- ε. Κανονικοποίηση τιμών.

Οι εργασίες δ. και ε., κωδικοποιήθηκαν και εκτελούνται κατά την εφαρμογή του εκάστοτε μοντέλου (η κωδικοποίηση τους περιγράφεται αναλυτικά στο 5^ο Κεφάλαιο).

4.2.1 Επιλογή Χαρακτηριστικών

Μία σημαντική κατηγορία εργασιών, που εντάσσεται στο πλαίσιο της προεπεξεργασίας, αφορά στη μείωση των δεδομένων. Ένα βασικό χαρακτηριστικό της εξόρυξης γνώσης από βάσεις δεδομένων είναι η απαίτηση δυνατότητας χειρισμού μεγάλου όγκου δεδομένων. Όμως, μεγάλος όγκος δεδομένων μπορεί να προκαλέσει προβλήματα στην επεξεργασία και μεγάλη καθυστέρηση στην υλοποίηση των αναλύσεων. Για αυτό το λόγο, η μείωση των δεδομένων είναι σε ορισμένες εφαρμογές απαραίτητη. Η μείωση των δεδομένων δεν αποτελεί μια τετριμμένη εργασία, καθότι τα αποτελέσματα που προκύπτουν από την ανάλυση των μειωμένων δεδομένων, θα πρέπει να είναι ίδια ή σχεδόν ίδια με εκείνα που θα πρόκυπταν από την ανάλυση του συνόλου των δεδομένων. Η *επιλογή των σημαντικών χαρακτηριστικών (feature selection)*, είναι μια εργασία μείωσης του όγκου. Τα διαθέσιμα δεδομένα, όπως ισχύει και στην περίπτωση της παρούσας εργασίας, αποτελούνται από εγγραφές (*γραμμές*) με πολλά στοιχεία - χαρακτηριστικά (*στήλες*). Ωστόσο, δεν είναι όλα τα χαρακτηριστικά χρήσιμα για όλες τις εργασίες εξόρυξης γνώσης. Πολλά χαρακτηριστικά περιέχουν πληροφορία η οποία δεν σχετίζεται με το αντικείμενο ανάλυσης. Επίσης, υπάρχει περίπτωση να καταγράφεται, σε διάφορα χαρακτηριστικά, διαφορετική εκδοχή της ίδιας πληροφορίας. Για κάθε εργασία που διεξάγεται, απαιτείται διερεύνηση για την εύρεση των κατάλληλων μεταβλητών. Σε ορισμένες περιπτώσεις, η ταυτόχρονη παρουσία ορισμένων χαρακτηριστικών – μεταβλητών οι οποίες χαρακτηρίζονται από έντονη αλληλεξάρτηση, ενδέχεται να είναι περιττή ή/και επιζήμια. Για τους παραπάνω λόγους, απαιτείται η προσεκτική επιλογή εκείνων των χαρακτηριστικών, τα οποία είναι τα πλέον κατάλληλα για συγκεκριμένη εργασία εξόρυξης γνώσης [28].

Στην περίπτωση της εφαρμογής της παρούσας ανάλυσης, η εξόρυξη γνώσης από τα δεδομένα που είναι διαθέσιμα, βασίζεται στην παρατήρηση προτύπων, δηλαδή στην αναζήτηση των αιτιών τα οποία οδηγούν σε συγκεκριμένο αποτέλεσμα, το οποίο είναι η εκδήλωση αστοχία της σήραγγας. Επομένως, τα χαρακτηριστικά τα οποία αφορούν καθαυτό στο αποτέλεσμα ή στην επίδραση αυτού (οικονομική ζημιά, ανθρώπινες απώλειες, χρονική καθυστέρηση), θεωρήθηκε ότι δεν έπρεπε να ληφθούν υπόψη και για το λόγο αυτό τα χαρακτηριστικά “failure type”, “portal failure”, “incident - reports and causes”, “third party impact”, “fatalities”, “losses” και “delays” εξαιρέθηκαν.

Επιπλέον, από τις εγγραφές της βάσης, παρατηρήθηκε ότι η αστοχία σήραγγας μετά την ολοκλήρωση της κατασκευής της, δηλαδή κατά τη λειτουργία της, στην πλειοψηφία των

περιπτώσεων οφείλεται στην επίδραση τυχηματικών φορτίων (φωτιές, σεισμοί) και ως εκ τούτου εξαιρέθηκαν οι εγγραφές της βάσης δεδομένων που αφορούν σε αστοχία κατά τη φάση κατασκευής. Εξαιρέθηκαν ακόμα τα χαρακτηριστικά “construction/operation” (όλες οι εγγραφές είχαν την τιμή “operation”), όπως επίσης και τα “year”, “country”, “project” και “source of information” τα οποία αφορούν απλώς στη καταγραφή του περιστατικού.

Όσον αφορά στα υπόλοιπα χαρακτηριστικά, τα οποία θεωρούνται σημαντικά, επισημαίνεται το εξής: οι αλγόριθμοι K-Means και K-Prototypes, στους οποίους βασίστηκε η ανάπτυξη των μοντέλων μας, δύναται να εφαρμοστούν μόνο σε εγγραφές που περιλαμβάνουν τιμές για όλα τα χαρακτηριστικά – γνωρίσματα. Επομένως το τμήμα της βάσης δεδομένων που τελικώς θα αναλυθεί, θα πρέπει να διαθέτει στήλες με ποσοστό πλήρωσης εκατό τοις εκατό (100%). Επομένως, τα χαρακτηριστικά “earthquake seismicity”, “geological information”, “stress conditions” και “water conditions”, παρά το γεγονός ότι σχετίζονται απολύτως με τα αίτια αστοχίας, δεδομένης της ιδιαίτερα χαμηλής πλήρωσης των στηλών (0%, 27%, 0% και 2% αντιστοίχως), θεωρήθηκε ότι θα ήταν αδύνατη η πλήρωση των χαμένων τιμών στο χρονικό πλαίσιο εκπόνησης της εργασίας και επομένως εξαιρέθηκαν.

Κατόπιν των ανωτέρω, επιλέχθηκαν ως σημαντικά χαρακτηριστικά τα “tunnel use”, “length”, “diameter”, “cutting technique”, “overburden” και “rock/soil”. Λαμβάνοντας όμως υπόψη μόνο τα σημαντικά χαρακτηριστικά, προέκυψαν εβδομήντα δύο (72) εγγραφές για τις οποίες δεν υπήρχαν τιμές για κανένα από αυτά τα χαρακτηριστικά, επομένως προέκυψε ένα εκμεταλλεύσιμο τμήμα της βάσης δεδομένων, το οποίο αποτελείται από τριακόσιες πέντε (305) εγγραφές, με την αντίστοιχη πλήρωση χαρακτηριστικών που φαίνεται στον Πίνακα 2.

<i>Στοιχείο – Χαρακτηριστικό</i>	<i>Ποσοστό Πλήρωσης</i>
Tunnel Use (Χρήση)	71 %
Cutting Technique (Μέθοδος Διάνοιξης)	61 %
Length (Μήκος) (m)	57 %
Diameter (Διάμετρος) (m)	38 %
Overburden (Υπερκείμενα) (m)	36 %
Rock / Soil (Βράχος / Έδαφος)	60 %

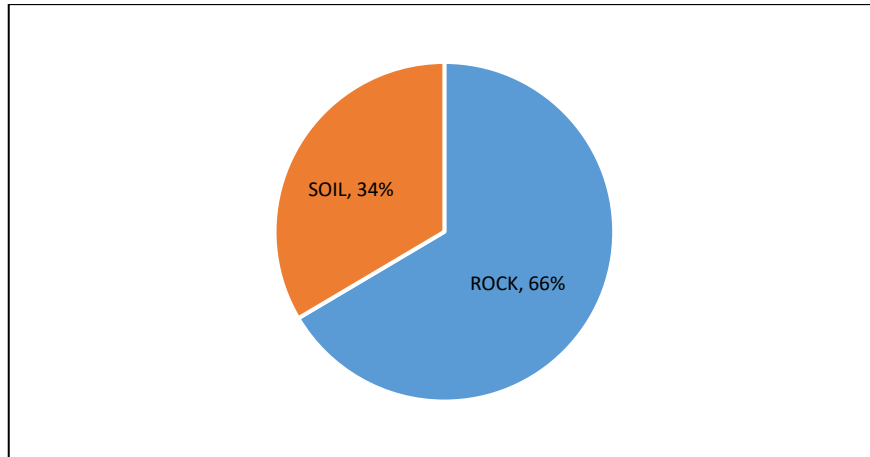
Πίνακας 2. Στοιχεία - χαρακτηριστικά περιστατικών αστοχίας σηράγγων που συμπεριλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων και αντίστοιχα ποσοστά πλήρωσης.

4.2.2 Πλήρωση Χαμένων Τιμών

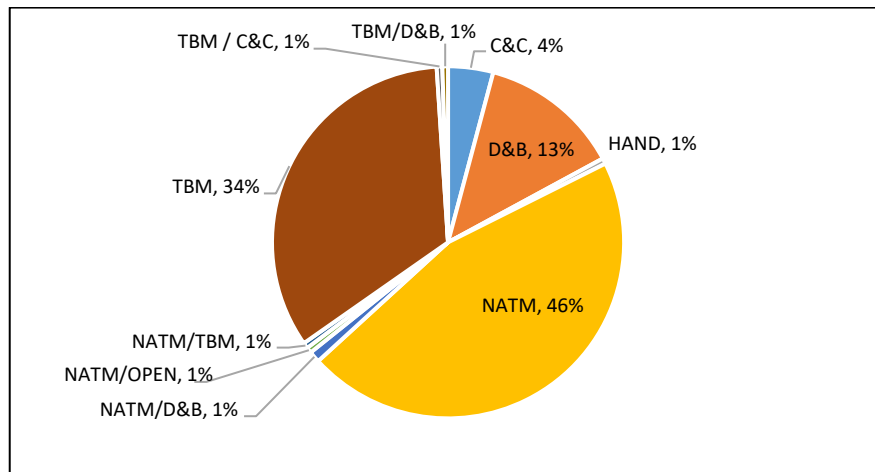
Η ύπαρξη *χαμένων τιμών* (*missing values*) είναι ένα σύνηθες πρόβλημα στα δεδομένα το οποίο μπορεί να προκύπτει για διάφορους λόγους [28]. Στην περίπτωση που εξετάστηκε οι χαμένες τιμές έχουν προκύψει καθότι η πληροφορία που διατίθεται στο ευρύ κοινό, για τις αστοχίες των σηράγγων, αλλά και σε ορισμένες περιπτώσεις για στοιχεία που αφορούν στην κατασκευή των σηράγγων, δεν είναι επαρκής. Επομένως, η εύρεση στοιχείων πραγματοποιείται μέσω αναζήτησης στη διεθνή βιβλιογραφία, όπου καταγράφονται συγκεκριμένα περιστατικά αστοχιών ή δημοσιεύονται αντίστοιχοι κατάλογοι και μέσω αναζήτησης στο διαδίκτυο για τυχόν δημοσιευμένα σχετικά άρθρα. Λαμβάνοντας υπόψη όμως το γεγονός ότι η δημιουργία της βάσης δεδομένων που παρασχέθηκε για την εκπόνηση της εργασίας, πραγματοποιήθηκε με τον προαναφερθέν τρόπο, αλλά και το ότι η τελευταία ενημέρωση της πραγματοποιήθηκε το έτος 2019, η δυνατότητα πλήρωσης των χαμένων τιμών ήταν εξαιρετικά μικρή. Πάραυτα, επιτεύχθηκε σε ένα χαμηλό βαθμό η αύξηση των ποσοστών πλήρωσης ανά χαρακτηριστικό και η προσθήκη νέων εγγραφών. Επομένως, το εκμεταλλεύσιμο τμήμα της βάσης δεδομένων, μετά την πλήρωση των χαμένων τιμών και με την προσθήκη των νέων, έφτασε τις τριακόσιες οχτώ (308) εγγραφές, με τα ποσοστά πλήρωσης χαρακτηριστικών που φαίνονται στον Πίνακα 3. Στα Διαγράμματα 1, 2, 3 και 4 φαίνονται ορισμένα στατιστικά στοιχεία που προκύπτουν από τα περιστατικά που συμπεριλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων και αφορούν αντιστοίχως τον τύπο του γεωυλικού (“rock/soil”) εντός του οποίου κατασκευάστηκαν οι σήραγγες, τη μέθοδο κατασκευής που χρησιμοποιήθηκε και τον τύπο της εμφανιζόμενης αστοχίας.

<i>Στοιχείο – Χαρακτηριστικό</i>	<i>Ποσοστό Πλήρωσης</i>
Tunnel Use (Χρήση)	73 % (+2%)
Cutting Technique (Μέθοδος Διάνοιξης)	63 % (+2%)
Length (Μήκος) (m)	51 % (+6%)
Diameter (Διάμετρος) (m)	43 % (+5%)
Overburden (Υπερκείμενα) (m)	43 % (+7%)
Rock / Soil (Βράχος / Έδαφος)	62 % (+2%)

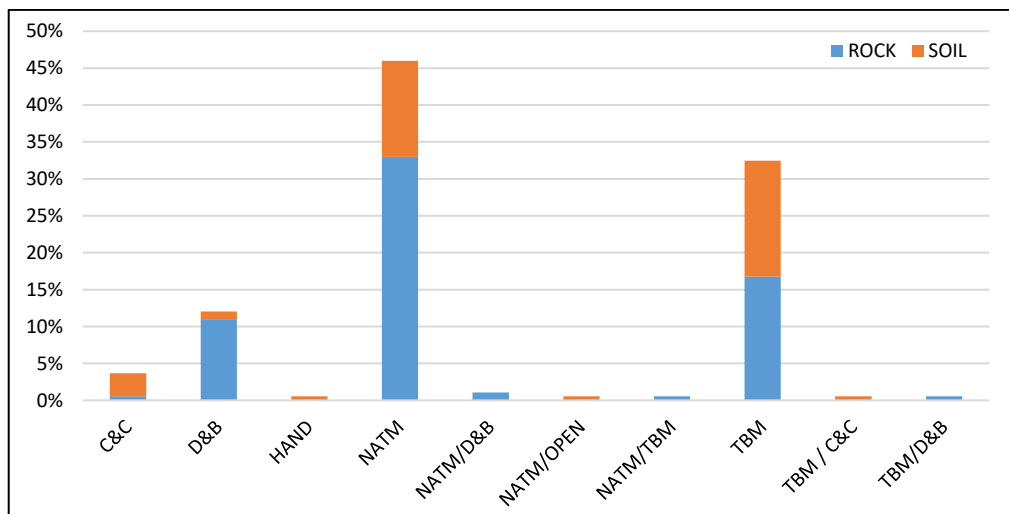
Πίνακας 3. Χαρακτηριστικά περιστατικών αστοχίας σηράγγων που συμπεριλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων, μετά την πλήρωση των χαμένων τιμών, και αντίστοιχα ποσοστά πλήρωσης.



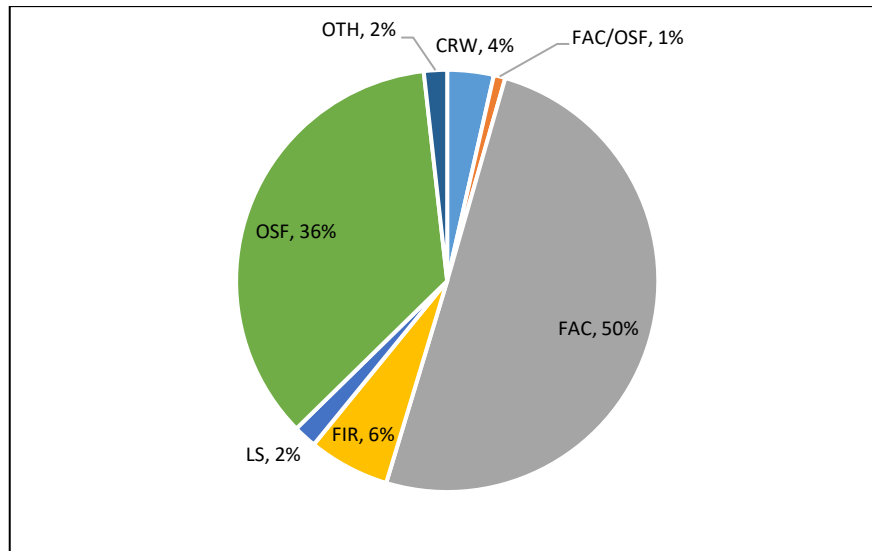
Διάγραμμα 1. Γεωολικό εντός του οποίου κατασκευάστηκαν οι σήραγγες που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων.



Διάγραμμα 2. Μέθοδος κατασκευής σηράγγων που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων.



Διάγραμμα 3. Γεωολικό σε συσχέτιση με τη μέθοδο κατασκευής των σηράγγων που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων.



Διάγραμμα 4. Τύποι αστοχίας των σηράγγων που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων (FAC: Face Collapse, OSF: Overstressing/Support Failure, FIR: Fire, CRW: Crown Wedge Failure, LS: Lateral Support Failure, OTH: Other).

Από το Διάγραμμα 4 δύναται να συναχθεί το συμπέρασμα ότι η αστάθεια μετώπου εκσκαφής (FAC: Face Collapse), αποτελεί με σχετικά μεγάλη διαφορά τον πιο συχνά εμφανιζόμενο τύπο αστοχίας. Οι παραμορφώσεις ενός τμήματος της σήραγγας, αρχίζουν να αναπτύσσονται σε κάποια απόσταση μπροστά από το μέτωπο εκσκαφής. Εάν αυτές οι παραμορφώσεις υπολογιστούν λανθασμένα και δεν μετριάστουν έγκαιρα αλλά και σωστά (εφαρμόζοντας τις κατάλληλες διορθωτικές τεχνικές), αναπτύσσονται ανεξέλεγκτα και οδηγούν σε αστάθεια και αστοχία, αναλόγως της αντοχής και της παραμορφωσιμότητας του γεωυλικού. Μια αρχική αξιολόγηση των συνθηκών ευστάθειας του μετώπου, μπορεί να είναι απαραίτητη για την αποτροπή επερχόμενων αστοχιών και αποτελεί βασικό παράγοντα επιτυχίας για τα έργα κατασκευής σηράγγων, ειδικά για ρηχές σήραγγες (όπως οι περισσότερες από τις αστικές σήραγγες του μετρό) [5]. Οι τύπος αστοχίας υποστήριξης ως αποτέλεσμα ανάπτυξης υπερβολικής τάσης/ανεπαρκούς τοποθέτησης μέτρων υποστήριξης (OSF: Overstressing/Support Failure), είναι αυτός που εμφανίζεται σε σχεδόν το ένα τρίτο των περιστατικών. Είναι υψίστης σημασίας να τοποθετηθούν τα κατάλληλα μέτρα υποστήριξης με την ενδεδειγμένη ακολουθία εγκατάστασης και την κατάλληλη χρονική στιγμή, προκειμένου να αποφευχθούν πιθανές αστοχίες [5].

4.2.3 Διαχείριση Θορυβωδών Δεδομένων

Ως θορυβώδη χαρακτηρίζονται τα δεδομένα που περιέχουν τιμές - εξαιρέσεις ή εσφαλμένες τιμές, τα οποία δεν προσφέρουν πληροφορία χρήσιμη για την ανάλυση. Η ύπαρξη θορύβου πρέπει να αντιμετωπίζεται κατά την προεπεξεργασία των δεδομένων καθώς προκαλεί προβλήματα στην εφαρμογή των αλγορίθμων εξόρυξης [28]. Στο τελικώς διαμορφωθέν εκμεταλλεύσιμο τμήμα της βάσης δεδομένων, δεν βρέθηκαν τιμές - εξαιρέσεις, βρέθηκαν όμως έξι (6) εσφαλμένες τιμές οι οποίες αντικαταστάθηκαν με τις ορθές.

4.2.4 Μετατροπή Ονομαστικών Τιμών σε Αριθμητικές

Όπως προαναφέρθηκε στο Κεφάλαιο 1.4, στην παρούσα εργασία περιγράφεται η ανάπτυξη δύο (2) μοντέλων. Το πρώτο μοντέλο (Tunnel Failure Model 1 - TFM1) βασίζεται στον αλγόριθμο συσταδοποίησης K-Means και το δεύτερο μοντέλο (Tunnel Failure Model 2 - TFM2) στον αλγόριθμο συσταδοποίησης K-Prototypes. Η εφαρμογή όμως του αλγόριθμου K-Means, προϋποθέτει την ύπαρξη μόνο αριθμητικών γνωρισμάτων, καθώς βασίζεται στον υπολογισμό της τετραγωνικής ευκλείδειας απόστασης των σημείων – εγγραφών (βλ. Κεφ. 3.3.3). Στην περίπτωση της παρούσας ανάλυσης, τα σημαντικά χαρακτηριστικά ήταν τρία αριθμητικά (“length”, “diameter”, “overburden”) και τρία ονομαστικά – περιγραφικά (“tunnel use”, “cutting technique”, “rock/soil”). Επομένως, οι τιμές των περιγραφικών μεταβλητών έπρεπε να μετατραπούν σε αριθμητικές προκειμένου να καταστεί δυνατός ο υπολογισμός των τετραγωνικών ευκλείδειων αποστάσεων. Υπενθυμίζεται ότι για την εφαρμογή του αλγορίθμου K-Prototypes δεν απαιτείται αυτή η μετατροπή, διότι για τα περιγραφικά χαρακτηριστικά γίνεται υπολογισμός του μέτρου ανομοιότητας (βλ. Κεφ. 3.3.4.1). Στον Πίνακα 4 φαίνονται οι αριθμητικές τιμές (*numerical values*) που ορίστηκαν για κάθε πιθανή περιγραφική τιμή (*descriptive values*) που δύναται να λάβουν τα περιγραφικά γνωρίσματα – μεταβλητές που ελήφθησαν υπόψη. Η εργασία καθαυτή της μετατροπής των τιμών έχει κωδικοποιηθεί και εκτελείται, σύμφωνα με τις τιμές που ορίστηκαν, κατά την εφαρμογή του 1^{ου} Μοντέλου (TFM1).

<i>Tunnel Use</i>		<i>Cutting Technique</i>		<i>Rock / Soil</i>	
<i>Descriptive Value</i>	<i>Numerical Value</i>	<i>Descriptive Value</i>	<i>Numerical Value</i>	<i>Descriptive Value</i>	<i>Numerical Value</i>
Sewer	1	Hand	1	Rock	1
Ventilation	2	C&C	2	Soil	2
Hydro	3	D&B	3		
Rail	4	NATM/D&B	4		
Train (lightrail)	5	NATM/Open	5		
Metro	6	NATM	6		
Road	7	NATM/TBM	7		
Walk	8	TBM/D&B	8		
		TBM/C&C	9		
		TBM	10		

Πίνακας 4. Μετατροπή ονομαστικών - περιγραφικών τιμών σε αριθμητικές (για την εφαρμογή του αλγόριθμου K-Means).

4.2.5 Κανονικοποίηση Τιμών

Ως *κανονικοποίηση (normalization)* ορίζεται η διαδικασία αντικατάστασης αριθμητικών τιμών που αφορούν σε γνωρίσματα – χαρακτηριστικά των εγγραφών μιας βάσης δεδομένων, με άλλες “καταλληλότερες” αριθμητικές τιμές. Η κανονικοποίηση γίνεται προκειμένου να καταστεί δυνατή η εφαρμογή ορισμένων μεθόδων εξόρυξης. Για παράδειγμα, στα Νευρωνικά Δίκτυα όταν οι τιμές κυμαίνονται από μηδέν (0) έως ένα (1) παρατηρείται καλύτερη λειτουργία. Επίσης, η μέθοδος K-Nearest Neighbors αλλά και οι μέθοδοι K-Means και K-Prototypes που εφαρμόζονται στην παρούσα εργασία, στις οποίες υπολογίζονται αποστάσεις μεταξύ των παρατηρήσεων, αντιμετωπίζουν πρόβλημα όταν κάποιες μεταβλητές εισόδου έχουν χαμηλές τιμές και άλλες μεταβλητές υψηλές τιμές. Το πρόβλημα που προκύπτει είναι ότι οι μεταβλητές με υψηλές τιμές είναι αυτές που ορίζουν την απόσταση των σημείων – παρατηρήσεων. Οι μεταβλητές με χαμηλές τιμές επηρεάζουν ελάχιστα την απόσταση και τελικά δεν επηρεάζουν σημαντικά τον υπολογισμό του αποτελέσματος [28].

Μία από τις πλέον χρησιμοποιούμενες μεθόδους κανονικοποίησης, είναι η *κανονικοποίηση ελάχιστου – μέγιστου (min-max normalization)*. Με τη μέθοδο αυτή οι αριθμητικές τιμές μετασχηματίζονται, έτσι ώστε να κυμαίνονται εντός ενός προκαθορισμένου εύρους τιμών. Η μέθοδος μετασχηματισμού είναι ο γραμμικός μετασχηματισμός. Για μία μεταβλητή A , με μεγαλύτερη παρατηρούμενη τιμή max_A και μικρότερη τιμή min_A , όλες οι ενδιάμεσες τιμές μπορούν να αντιστοιχηθούν σε άλλες τιμές, με εύρος που ορίζεται από τις νέες τιμές new_min_A και new_max_A , σύμφωνα με την **Εξίσωση 6** [28]:

$$x' = \frac{x - min_A}{max_A - min_A} (new_max_A - new_min_A) + new_min_A$$

Εξίσωση 6

, όπου x η τιμή της μεταβλητής A και x' η νέα τιμή. Το πλεονέκτημα αυτής της μεθόδου είναι ότι μπορεί να προκαθοριστεί ένα νέο εύρος τιμών για τα δεδομένα, με κατάλληλα για την εφαρμογή που μας ενδιαφέρει όρια, διατηρώντας την αναλογία των τιμών που υπήρχε στα αρχικά δεδομένα.

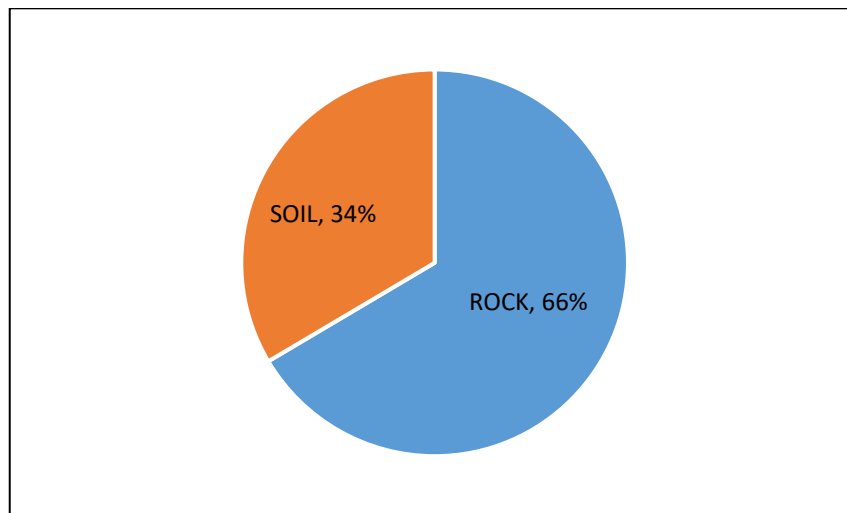
Εφόσον υπάρχει η δυνατότητα, η προαναφερθείσα εργασία, η οποία είναι απαραίτητο να εκτελεστεί προκειμένου να καταστεί η δυνατή η εφαρμογή των μοντέλων της παρούσας εργασίας, κωδικοποιήθηκε και εκτελείτε με την εφαρμογή του εκάστοτε μοντέλου. Το νέο εύρος τιμών προκαθορίστηκε μεταξύ των τιμών $min_A = 0$ και $max_A = 1$, ώστε οι τιμές όλων των χαρακτηριστικών κατά την ανάλυση των δεδομένων να κυμαίνονται μεταξύ 0 και 1, και να μην δίνεται μεγαλύτερη βαρύτητα σε κάποιο χαρακτηριστικό, λόγω υπολογισμού μεγαλύτερης τετραγωνικής ευκλείδειας απόστασης.

4.3 Dataset Εφαρμογής Μοντέλων (TFD_21)

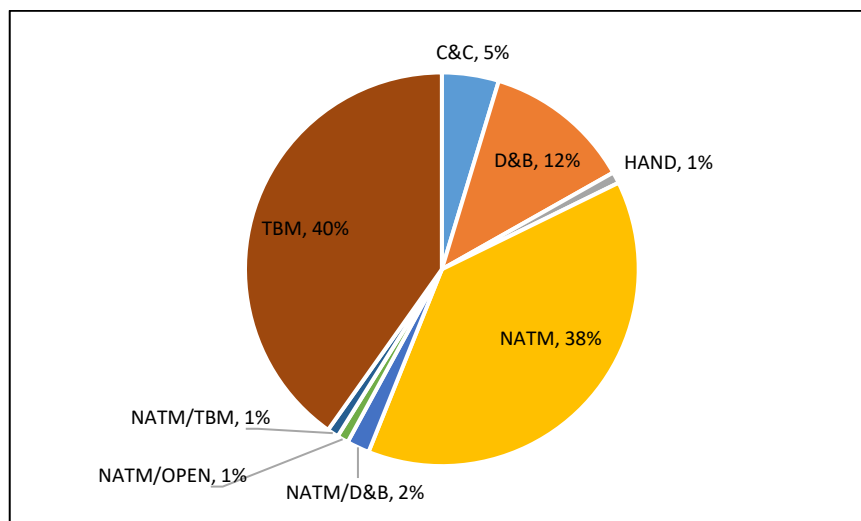
Κατόπιν ολοκλήρωσης των εργασιών προεπεξεργασίας των δεδομένων, προέκυψε ένα εκμεταλλεύσιμο τμήμα της βάσης δεδομένων το οποίο όμως, όπως προαναφέρθηκε, χαρακτηρίζεται από συγκεκριμένα ποσοστά πλήρωσης των στηλών που αντιστοιχούν στα σημαντικά γνωρίσματα, τα οποία θα χρησιμοποιηθούν προς εφαρμογή των μοντέλων. Οι αλγόριθμοι συσταδοποίησης K-Means και K-Prototypes όμως πραγματοποιούν υπολογισμό της τετραγωνικής ευκλείδειας απόστασης των σημείων – εγγραφών (βλ. **Κεφ. 3.3.3**). ή υπολογισμό του μέτρου ανομοιότητας (βλ. **Κεφ. 3.3.4.1**) και ως εκ τούτου απαιτείται η

ύπαρξη τιμών για όλα τα χαρακτηριστικά – γνωρίσματα της κάθε εγγραφής – σημείου, καθότι οι τιμές αυτές είναι οι συντεταγμένες των σημείων, οι οποίες θα χρησιμοποιηθούν για τους υπολογισμούς.

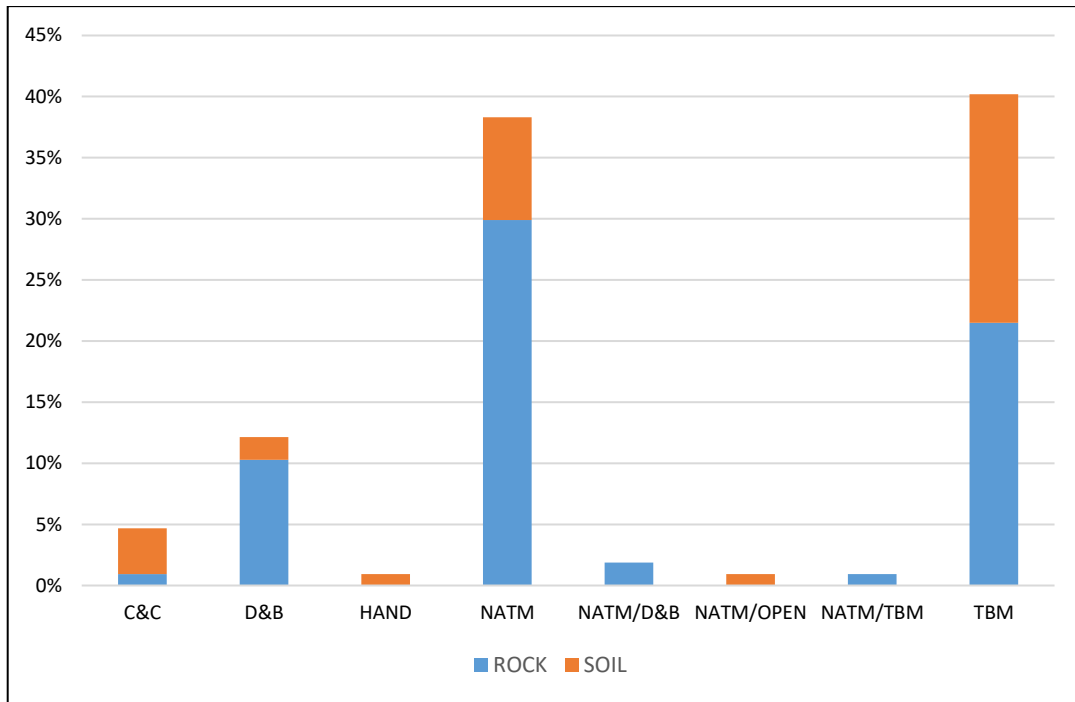
Επομένως, διαμορφώνεται το εκμεταλλεύσιμο τμήμα καταλλήλως, ώστε να προκύψει το απαιτούμενο dataset, το οποίο περιλαμβάνει εκατόν επτά (107) εγγραφές – περιστατικά αστοχίας. Στα Διαγράμματα 5, 6, 7 και 8 φαίνονται ορισμένα στατιστικά στοιχεία που προκύπτουν από τα περιστατικά που συμπεριλαμβάνονται στο dataset εφαρμογής των μοντέλων (Tunnel Failure Dataset 2021, TFD_21) και αφορούν αντιστοίχως στο γεωυλικό εντός του οποίου κατασκευάστηκαν οι σήραγγες, τη μέθοδο κατασκευής καθαυτή αλλά και την ίδια σε συνάρτηση με το γεωυλικό και τον τύπο αστοχίας.



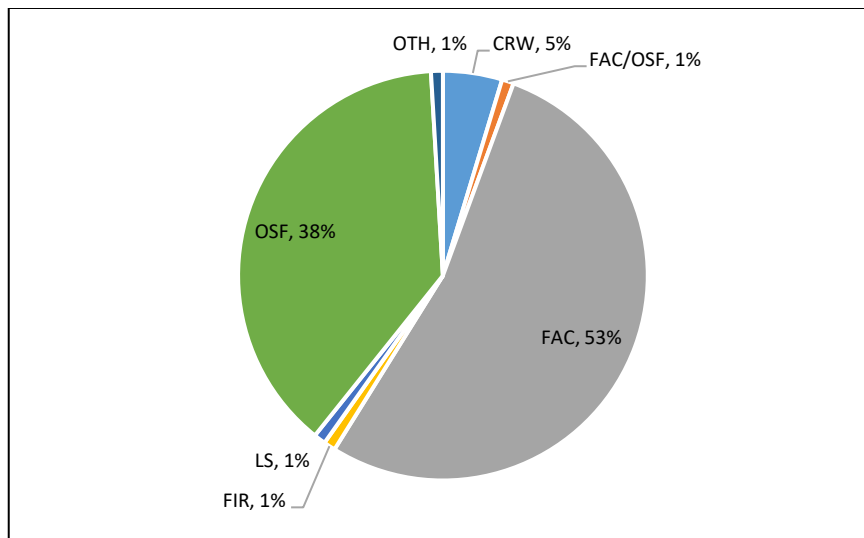
Διάγραμμα 5. Γεωυλικό εντός του οποίου κατασκευάστηκαν οι σήραγγες που περιλαμβάνονται στο dataset TFD_21.



Διάγραμμα 6. Μέθοδος κατασκευής των σηράγγων που περιλαμβάνονται στο dataset TFD_21.



Διάγραμμα 7. Γεωολικό σε συσχέτιση με τη μέθοδο κατασκευής των σηράγγων που περιλαμβάνονται στο dataset TFD_21.



Διάγραμμα 8. Τύποι αστοχίας των σηράγγων που περιλαμβάνονται στο dataset TFD_21 (FAC: Face Collapse, OSF: Overstressing/Support Failure, FIR: Fire, CRW: Crown Wedge Failure, LS: Lateral Support Failure, OTH: Other).

4.4 Ανάλυση Κυρίων Συνιστωσών (Principal Component Analysis)

Ένας σημαντικός προβληματισμός που προκύπτει όσον αφορά στη διαμόρφωση του dataset TFD_21, επί του οποίου θα εφαρμοστούν οι αλγόριθμοι, είναι το αν αυτό αποτελεί αντιπροσωπευτικό τμήμα της βάσης δεδομένων, το οποίο διατηρεί τα αρχικά μοτίβα συσχετίσεων μεταξύ των χαρακτηριστικών. Για την εξέταση ισχύος αυτής της αναγκαίας συνθήκης, εφαρμόστηκε η μέθοδος *Ανάλυσης Κυρίων Συνιστωσών (Principal Component Analysis, PCA)*, με σκοπό την αξιοποίηση των *παραγόντων βαρύτητας (loadings)* των αρχικών μεταβλητών προς εξαγωγή των σημαντικών χαρακτηριστικών, για τα οποία στη συνέχεια θα εξεταζόταν η διατήρηση των συσχετίσεων.

Η PCA αποτελεί την πιο διαδεδομένη μέθοδο ανάλυσης πολλών μεταβλητών, που ως στόχο έχει την μείωση της διαστατικότητας των δεδομένων, μέσω ανεύρεσης ορισμένων νέων μεταβλητών, οι οποίες είναι γραμμικοί συνδυασμοί των αρχικών μεταβλητών και δεν συσχετίζονται μεταξύ τους. Οι νέες αυτές μεταβλητές έχουν την δυνατότητα να εμπεριέχουν μεγάλο μέρος της ολικής μεταβλητότητας που αναπτύσσεται μεταξύ των αρχικών μεταβλητών, το οποίο κατανέμεται σε λιγότερες νέες μεταβλητές. Έτσι, το μεγαλύτερο μέρος της πληροφορίας που θα εξαγόταν αν λαμβάνονταν υπόψη οι αρχικές μεταβλητές, δύναται να εξαχθεί και από τις νέες μεταβλητές, οι οποίες είναι πλέον λιγότερες και η ανάλυση των οποίων, επομένως, είναι λιγότερο απαιτητική [29]. Η PCA είναι κατά βάση μέθοδος μείωσης των διαστάσεων και όχι μέθοδος επιλογής σημαντικών χαρακτηριστικών, τουλάχιστον κατά την έννοια επιλογής ορισμένων αρχικών διαστάσεων και απόρριψης των υπόλοιπων. Ωστόσο, παραδείγματα εφαρμογής της PCA για την επιλογή χαρακτηριστικών έχουν δημοσιευτεί από τους Shie et al. (2012) [30] και τον Chen (2011) [31], όπου αξιοποιήθηκαν οι παράγοντες βαρύτητας των αρχικών μεταβλητών (loadings) στις κύριες συνιστώσες.

4.4.1 Περιγραφή Μεθόδου

Έστω ένα σύνολο δεδομένων με k γραμμές και n γνωρίσματα (ή διαστάσεις - στήλες). Με την PCA αναπτύσσεται ένα σύστημα m κάθετων διανυσμάτων, όπου $m < n$, και προβάλλονται τα δεδομένα σε ένα νέο χώρο m διαστάσεων. Με αυτόν τον τρόπο δημιουργούνται γραμμικοί συνδυασμοί των αρχικών μεταβλητών οι οποίοι είναι

ασυσχετίστοι μεταξύ τους και περιέχουν το μεγαλύτερο μέρος της διακύμανσης των αρχικών μεταβλητών [28].

Η μέθοδος βασίζεται στην ακόλουθη διαδικασία ανάλυσης:

α. Από n μεταβλητές (X_1, X_2, \dots, X_n), δημιουργούνται n συνδυασμοί αυτών (Z_1, Z_2, \dots, Z_n), με τρόπο ώστε να μη συσχετίζονται μεταξύ τους. Για να μην συσχετίζονται, θα πρέπει οι μεταβλητές Z_i να μετρούν διαφορετικές “διαστάσεις” των στοιχείων, δηλαδή διαφορετικά γνωρίσματα - χαρακτηριστικά.

β. Οι διακυμάνσεις (μεταβλητότητες) που αναπτύσσονται μεταξύ των Z_i μεταβλητών, διαβαθμίζονται με κατάλληλο τρόπο ώστε η Z_1 μεταβλητή να εξηγεί όσο το δυνατόν μεγαλύτερο ποσοστό της ολικής μεταβλητότητας, η Z_2 το δεύτερο μεγαλύτερο κ.ο.κ., τηρούμενης πάντα της σχέσης: $\lambda_1 > \lambda_2 > \dots > \lambda_n$, όπου λ_i η διακύμανση. Οι νέες μεταβλητές Z_i , αποτελούν τις κύριες συνιστώσες. Με αυτόν τον τρόπο δημιουργούνται ολιγάριθμες Z συνιστώσες, οι οποίες εμπεριέχουν μεγάλο ποσοστό της ολικής διακύμανσης $\Sigma \lambda_i$. Παράλληλα, δημιουργούνται δευτερεύουσες συνιστώσες, οι οποίες εμπεριέχουν πολύ μικρό ποσοστό της ολικής διακύμανσης και επομένως το στατιστικό αποτέλεσμα αυτών μπορεί να αγνοηθεί, χωρίς ουσιαστική απώλεια πληροφορίας. Βάση υπολογισμών της PCE είναι η *μήτρα συσχετίσεων (correlation matrix)* των μεταβλητών. Η εφαρμογή της μεθόδου θεωρείται επιτυχής, δηλαδή θεωρείται ότι παρέχεται ουσιωδώς πληροφορία, όταν κάποιοι συντελεστές συσχέτισης είναι, κατ’ απόλυτη τιμή, υψηλοί (π.χ. $r \geq \pm 0,700$). Υπό την προϋπόθεση αυτή, ένα σύνολο αρχικών μεταβλητών μπορεί να αντιπροσωπεύεται από δύο ή τρεις κύριες συνιστώσες. Στην περίπτωση που προκύπτουν αρχικές μεταβλητές με πολύ υψηλές τιμές συσχετίσεων ($r > \pm 0,990$), αυτές θα πρέπει να θεωρούνται πλεονάζουσες και κάποιες να απορρίπτονται πριν την εφαρμογή της μεθόδου [29].

Τα στάδια της PCA έχουν ως ακολούθως:

α. Τυποποίηση των αρχικών μεταβλητών X_1, X_2, \dots, X_n σύμφωνα με την **Εξίσωση 7**, έτσι ώστε να προκύπτει μέσος όρος 0 και διακύμανση ίση με 1:

$$X'_i = \frac{X_i - \bar{X}}{s}$$

Εξίσωση 7

β. Στην συνέχεια υπολογίζονται οι *συνδιασπορές (covariance)* των μεταβλητών σύμφωνα με την **Εξίσωση 8** και προκύπτει ο *πίνακας συνδιασποράς (covariance matrix)* σύμφωνα με την **Εξίσωση 9**. Η συνδιασπορά αποτελεί μέτρο για το πως μεταβάλλονται μία μεταβλητή σε σχέση με μία άλλη. Ένα η τιμή της είναι θετική, τότε οι δύο μεταβλητές αυξάνονται ταυτόχρονα, εάν όμως είναι αρνητική, τότε η αύξηση της μίας συνεπάγεται μείωση της άλλης [28]. Ο πίνακας συνδιασποράς λόγω τυποποίησης των αρχικών μεταβλητών μεταπίπτει στη *μήτρα συσχετίσεων (correlation matrix)* των μεταβλητών, η οποία έχει τη μορφή του πίνακα C.

$$c(X_i, X_j) = \frac{\sum_{y=1}^k (X_{iy} - \bar{X}_i)(X_{jy} - \bar{X}_j)}{k - 1}$$

Εξίσωση 8

, όπου k το σύνολο των παρατηρήσεων (γραμμές) που αφορούν στα γνωρίσματα – μεταβλητές X_i και X_j .

$$C = \begin{pmatrix} 1 & c_{12} & \dots & c_{1n} \\ c_{21} & 1 & & c_{2n} \\ \dots & & 1 & \\ c_{n1} & c_{n2} & & 1 \end{pmatrix}$$

Εξίσωση 9

γ. Υπολογισμός των ιδιοδιανυσμάτων του πίνακα συνδιασποράς. Τα ιδιοδιανύσματα αυτά έχουν μήκος ίσο με ένα. Τα ιδιοδιανύσματα ονομάζονται *κύριες συνιστώσες (principal components)* και τα αρχικά δεδομένα μπορούν να εκφραστούν ως γραμμικοί συνδυασμοί των κυρίων συνιστωσών. Δηλαδή, οι βασικές συνιστώσες μπορούν να αποτελέσουν ένα νέο σύστημα αξόνων για τα δεδομένα. Η διαδικασία υπολογισμού τους είναι η ακόλουθη:

ι. Η πρώτη κύρια συνιστώσα Z_1 προκύπτει από το γραμμικό συνδυασμό των n αρχικών μεταβλητών.

$$Z_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \dots + \alpha_{1n}X_n$$

Εξίσωση 10

, όπου α_{ij} ειδικός συντελεστής στάθμισης (*weight*) της μεταβλητής j στην i συνιστώσα, με τον περιορισμό ότι:

$$\alpha_{11}^2 + \alpha_{12}^2 + \dots + \alpha_{1n}^2 = 1$$

Εξίσωση 11

Από τον ανωτέρω περιορισμό εξασφαλίζεται η εκτίμηση της μέγιστης διακύμανσης λ_1 της Z_1 .

ii. Η δεύτερη κύρια συνιστώσα προκύπτει ομοίως ως:

$$Z_2 = \alpha_{21}X_1 + \alpha_{22}X_2 + \dots + \alpha_{2n}X_n$$

Εξίσωση 12

, με τον περιορισμό ότι:

$$\alpha_{21}^2 + \alpha_{22}^2 + \dots + \alpha_{2n}^2 = 1$$

Εξίσωση 13

και με περιορισμό ότι ο συντελεστής συσχέτισης μεταξύ των κυρίων συνιστωσών Z_1 και Z_2 να είναι ίσος με μηδέν (0).

iii. Με τον ίδιο τρόπο υπολογίζονται και η τρίτη κύρια συνιστώσα Z_3 , ασυσχέτιστη με τις δύο προηγούμενες.

Με τον ανωτέρω τρόπο δημιουργούνται n συνιστώσες Z_i , ισάριθμες δηλαδή κύριες συνιστώσες με τις αρχικές μεταβλητές, για τις οποίες υπολογίζονται οι διακυμάνσεις λ_n , οι οποίες ονομάζονται *χαρακτηριστικές ρίζες* ή *ιδιοτιμές* (*eigenvalues*), είναι ισάριθμες με τις n συνιστώσες και για τις οποίες ισχύουν τα ακόλουθα:

$$\lambda_1 > \lambda_2 > \dots > \lambda_n$$

$$\lambda_1 + \lambda_2 + \dots + \lambda_n = c_{11} + c_{22} + \dots + c_{nn}$$

Εξίσωση 14

δ. Οι κύριες συνιστώσες ταξινομούνται με βάση τις ιδιοτιμές τους, οι οποίες εκφράζουν τη σημαντικότητα των ιδιοδιανυσμάτων, ως προς τη δυνατότητα αναπαράστασης των δεδομένων. Δηλαδή όσο μεγαλύτερη η ιδιοτιμή, τόσο περισσότερη πληροφορία σχετικά με τη διασπορά των δεδομένων περιλαμβάνει η συνιστώσα. Με τον

τρόπο αυτό μπορούν να αφαιρεθούν οι λιγότερο σημαντικές, οι συνιστώσες δηλαδή των οποίων οι ιδιοτιμές είναι μικρές και επομένως δεν φέρουν σημαντική πληροφορία.

ε. Τέλος, υπολογίζονται τα φορτία συνιστωσών (*loadings*), δηλαδή οι συσχετίσεις μεταξύ των αρχικών μεταβλητών και των κύριων συνιστωσών οι οποίες απεικονίζουν, την δράση που αναπτύσσουν οι αρχικές μεταβλητές στη δημιουργία των κυρίων συνιστωσών. Η δράση αυτή θεωρείται εξαιρετικά ισχυρή για τιμές κοντά στο $\pm 1,00$ (ισχυρή θετική ή αρνητική συσχέτιση) και ασήμαντη για τιμές κοντά στο μηδέν. Όσο υψηλότερα δηλαδή είναι τα φορτία, τόσο σημαντικότερες είναι οι μεταβλητές για το σχηματισμό των κύριων συνιστωσών. Τα φορτία υπολογίζονται από την **Εξίσωση 15**:

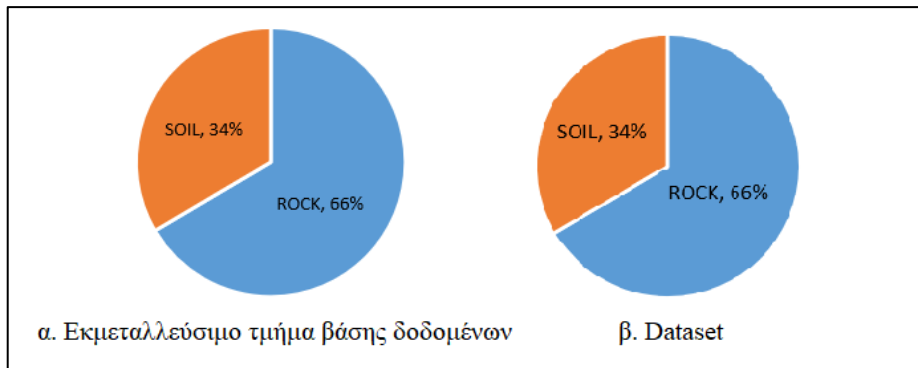
$$l_{ij} = \frac{\alpha_{ij}}{s_j} \sqrt{\lambda_i}$$

Εξίσωση 15

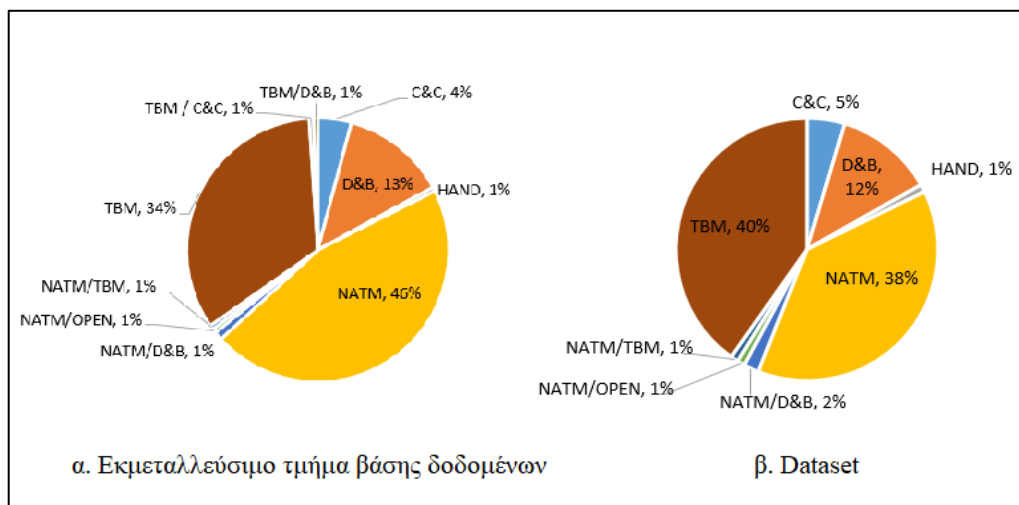
, όπου l_{ij} είναι το φορτίο της μεταβλητής j για την i συνιστώσα, α_{ij} είναι ο συντελεστής στάθμισης της μεταβλητής j για την i συνιστώσα επίσης, λ_i είναι η χαρακτηριστική ρίζα της i συνιστώσας και s_j είναι η τυπική απόκλιση της μεταβλητής j [29].

4.4.2 Εφαρμογή Μεθόδου - Αποτελέσματα

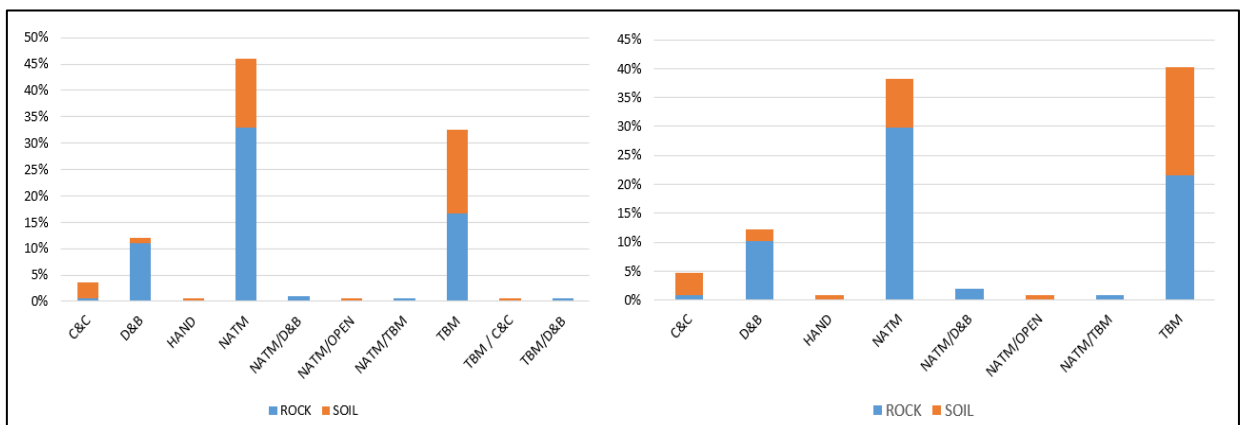
Η εφαρμογή της μεθόδου PCA στο dataset TFD_21, κωδικοποιήθηκε κάνοντας χρήση των βιβλιοθηκών της γλώσσας προγραμματισμού Python. Από την εφαρμογή της PCA, προέκυψαν ως “σημαντικότερα” χαρακτηριστικά για την δημιουργία των κυρίων συνιστωσών (pc1 και pc2) τα “rock/soil” και “cutting technique” με φορτία 0,97 και -0,89 αντίστοιχα. Ως εκ τούτου, θεωρήθηκε ότι ένας δείκτης δυνατότητας αντιπροσώπευσης του εκμεταλλεύσιμου τμήματος της βάσης δεδομένων μας από το dataset TFD_21, θα ήταν η διατήρηση των στατιστικών χαρακτηριστικών που αφορούν στις δύο αυτές μεταβλητές (“rock/soil” και “cutting technique”), αλλά και της συσχέτισης αυτών. Από την αντιπαράθεση των στατιστικών, η οποία απεικονίζεται και μέσω αντιπαράθεσης των Διαγραμμάτων 1-5, 2-4 και 3-7, στα Σχήματα 10, 11 και 12 που ακολουθούν, φαίνεται ότι διατηρούνται τα ζητούμενα χαρακτηριστικά σε ποσοστό 100% για το “rock/soil”, 92% για το “cutting technique” και 82% ως προς τη συσχέτιση αυτών. Επομένως το dataset TFD_21 θεωρήθηκε κατάλληλο για την εφαρμογή των μοντέλων.



Σχήμα 10. Γεωυλικό εντός του οποίου κατασκευάστηκαν οι σήραγγες που περιλαμβάνονται: α. στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων και β. στο dataset TFD_21.

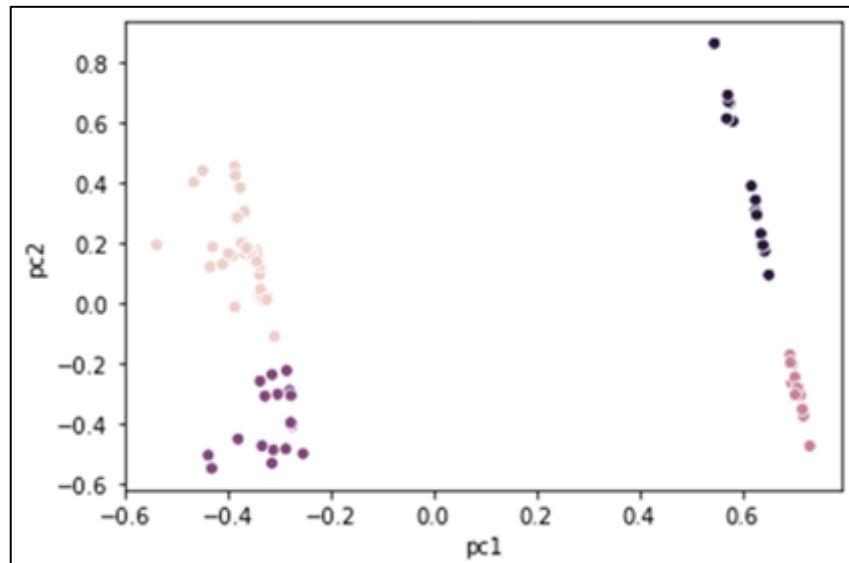


Σχήμα 11. Μέθοδος κατασκευής των σηράγγων που περιλαμβάνονται: α. στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων και β. στο dataset TFD_21.



Σχήμα 12. Γεωυλικό σε συσχέτιση με τη μέθοδο κατασκευής των σηράγγων που περιλαμβάνονται στο εκμεταλλεύσιμο τμήμα της βάσης δεδομένων (αριστερά) και στο dataset TFD_21 (δεξιά).

Ενδιαφέρον επίσης παρουσιάζει το **Διάγραμμα 9**, όπου απεικονίζονται τα δεδομένα που περιλαμβάνονται στο dataset TFD_21 σε δύο διαστάσεις, ως προς τις κύριες συνιστώσες pc1 και pc2.



Διάγραμμα 9. Απεικόνιση των δεδομένων που περιλαμβάνονται στο dataset TFD_21 σε δύο διαστάσεις, ως προς τις κύριες συνιστώσες pc1 και pc2, που προέκυψαν από την εφαρμογή της PCA. Εμφανής σχηματισμός συστάδων (clusters).

Ήδη από το διάγραμμα αυτό, πριν από την εφαρμογή των αλγορίθμων συσταδοποίησης, φαίνεται ότι είναι δυνατή η διαμόρφωση συστάδων (π.χ. τεσσάρων) που χαρακτηρίζονται από όμοια χαρακτηριστικά, δηλαδή όμοιες τιμές που αφορούν στις κύριες συνιστώσες που προέκυψαν από την εφαρμογή της PCA.

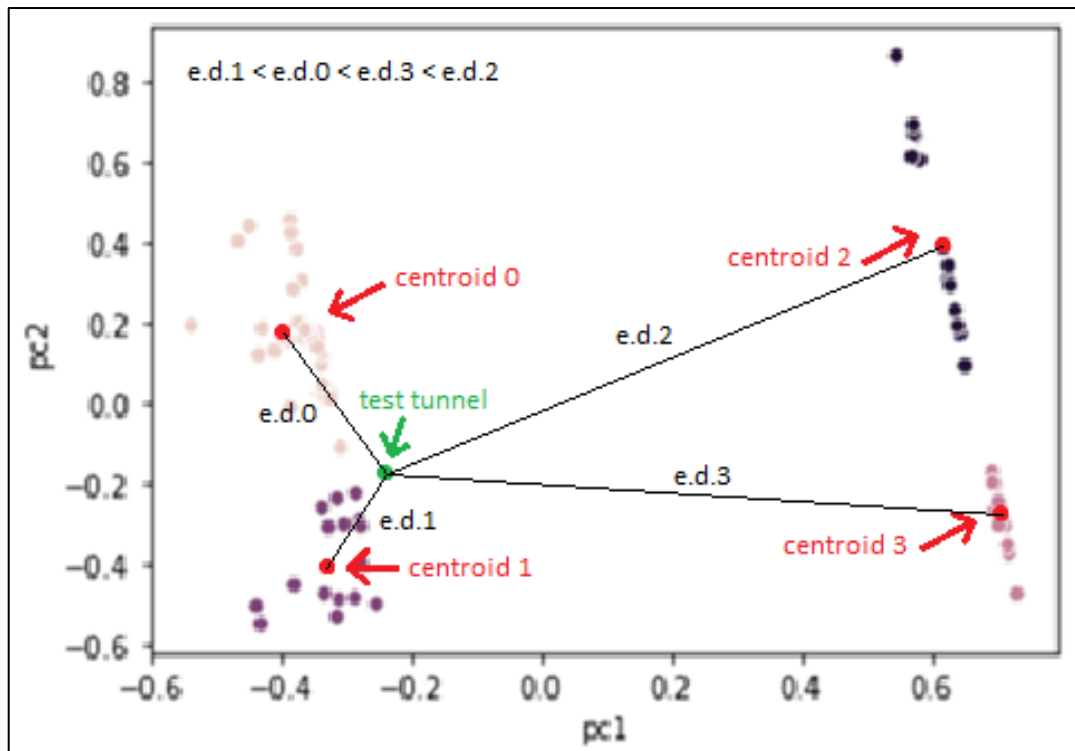
5 Μοντέλα Ανάλυσης Δεδομένων

Για την ανάλυση των δεδομένων που περιλαμβάνονται στο dataset TFD_21, προς τη διαμόρφωση συστάδων και την εκτίμηση πιθανών προτύπων (patterns), αναπτύχθηκαν δύο (2) μοντέλα. Το μοντέλο TFM1 (Tunnel Failure Model 1) βασίζεται στον αλγόριθμο K-Means και το μοντέλο TFM2 (Tunnel Failure Model 2) στον αλγόριθμο K-Prototypes. Τα μοντέλα ενσωματώνουν τους αλγόριθμους, προκειμένου να οριστούν συστάδες (clusters), αποτελούμενες από εγγραφές που παρουσιάζουν ομοιότητα ως προς τα γνωρίσματα – χαρακτηριστικά τους. Σε κάθε συστάδα (cluster), περιλαμβάνεται συγκεκριμένος αριθμός εγγραφών, δηλαδή περιστατικών αστοχίας σήραγγων. Ο αριθμός αυτός ορίζει την ποσοστιαία εμφάνιση του cluster στο dataset TFD_21 ή τη *σχετική συχνότητα εμφάνισης του cluster (relative frequency)*.

Ως περιβάλλον υλοποίησης, επιλέχθηκε η διαδικτυακή εφαρμογή, ελεύθερου λογισμικού και ανοιχτού κώδικα, Jupyter Notebook (version 6.0.3) και ως γλώσσα υλοποίησης η Python (version 3.8.3), με χρήση των βιβλιοθηκών pandas, seaborn, numpy, math, warnings, sklearn, matplotlib, spicy και mpl_toolkits.

Σκοπός ανάπτυξης των μοντέλων είναι η εκτίμηση προτύπων που οδηγούν σε συγκεκριμένη κατηγορία κινδύνου αστοχίας, αλλά και η εκτίμηση κινδύνου αστοχίας της εξεταζόμενης σήραγγας. Για την πραγματοποίηση της εξέτασης, προσδιορίζονται από τον χρήστη τα χαρακτηριστικά της σήραγγας ενδιαφέροντος και μέσω υπολογισμού της ευκλείδειας απόστασης από το κεντροειδές (centroid) του κάθε cluster, η σήραγγα χαρακτηρίζεται και κατηγοριοποιείται ως προς την κατηγορία κινδύνου στην οποία εμπίπτει. Συνεπώς, η εγγύτητα της υπό εξέταση σήραγγας προς το κεντροειδές (centroid) κάποιας συγκεκριμένης συσταδοποίησης (cluster), υποδεικνύει την ομοιότητα αυτής με τις εγγραφές που περιλαμβάνονται στο cluster. Στο Σχήμα 13 απεικονίζεται μέσω εφαρμογής PCA, σε δύο διαστάσεις, μία υπό εξέταση σήραγγα, η οποία παρουσιάζει μεγαλύτερη εγγύτητα με το cluster “1” (Euclidian distance $e.d.1 < e.d.0 < e.d.3 < e.d.2$) και η οποία συνεπώς φέρει μεγαλύτερη ομοιότητα με αυτό το cluster. Με βάση τα παραπάνω, αν μία υπό εξέταση σήραγγα ομοιάζει με το cluster που παρουσιάζει την υψηλότερη σχετική συχνότητα εμφάνισης στο dataset TFD_21, ομοιάζει με περισσότερες σήραγγες που παρουσίασαν αστοχία και θεωρείται ότι φέρει μεγαλύτερη πιθανότητα αστοχίας. Επομένως, λαμβάνοντας υπόψη τη σχετική συχνότητα αστοχίας του κάθε cluster, προκύπτουν κατ’αντιστοιχία οι κατηγορίες κινδύνου αστοχίας και η υπό εξέταση σήραγγα θεωρείται ότι

εμπίπτει στην κατηγορία κινδύνου, η οποία αντιστοιχίζεται μέσω της σχετικής συχνότητας αστοχίας στο cluster με το οποίο παρουσιάζει η ίδια τη μεγαλύτερη εγγύτητα (μικρότερη Ευκλείδεια απόσταση).



Σχήμα 13. Απεικόνιση μέσω εφαρμογής PCA, σε δύο διαστάσεις, υπό εξέταση σήραγγας (test tunnel) και εγγύτητας αυτής με τα κεντροειδή των συστάδων (Euclidian distance $e.d.1 < e.d.0 < e.d.3 < e.d.2$).

Η προαναφερθείσα θεώρηση είναι και η βάση εκτίμησης της σχετικής πιθανότητας αστοχίας της υπό εξέταση σήραγγας σε αντιπαράβολή με τα δεδομένα που προέρχονται από την χρησιμοποιούμενη βάση δεδομένων (dataset TFD_21). Δηλαδή, τα μοντέλα μπορούν να χρησιμοποιηθούν προκειμένου να γίνει σύγκριση σεναρίων κατασκευής σιράγγων και η αντιπαράβολή τους με συγκεκριμένα γνωρίσματα, ώστε να προκύψει το σενάριο που παρουσιάζει τον μικρότερο κίνδυνο αστοχίας και όχι ο πραγματικός κίνδυνος αστοχίας.

5.1 Μοντέλο TFM1

Το μοντέλο TFM1 αποτελείται από δύο τμήματα. Το πρώτο τμήμα ενσωματώνει τον αλγόριθμο K-Means, μέσω του οποίου διαμορφώνονται οι συστάδες (clusters) των περιστατικών αστοχίας σήραγγων που περιλαμβάνονται στο dataset TFD_21. Με βάση τα αποτελέσματα της συσταδοποίησης, γίνεται η σήμανση (labeling) των συστάδων η οποία αξιοποιείται και στο δεύτερο τμήμα του μοντέλου, όπου δίνεται τη δυνατότητα στο χρήστη να εισάγει τα χαρακτηριστικά μίας υπό εξέταση σήραγγας και να εξάγει αποτελέσματα που αφορούν στην ομοιότητά της με τα περιστατικά που περιλαμβάνονται στο TFD_21 και στον κίνδυνο αστοχίας της. Συγκεκριμένα, εξάγεται η πλησιέστερη περίπτωση αστοχίας σήραγγας (*κοντινότερος γείτονας – nearest neighbor*), ο βαθμός ομοιότητάς της με αυτή, με το cluster στο οποίο υπάγεται και με το σύνολο των περιστατικών του TFD_21, η σχετική πιθανότητα αστοχίας της και η κατηγορία κινδύνου αστοχίας στην οποία εμπίπτει.

5.1.1 Συσταδοποίηση (TFM1_A)

Αρχικά εισάγεται το προς ανάλυση dataset TFD_21 στο μοντέλο (Εικόνα 1).

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Thirre Tunnel	Road	NATM	5500	12.60	650.0	Rock
1	Lane cove Tunnel	Ventilation	NATM	130	8.00	17.0	Rock
2	Wienerwald tunnel, Vienna	Rail	NATM/D&B	13350	10.68	190.0	Rock
3	Lainzer railroad tunnel, Vienna	Rail	NATM	12800	5.55	26.0	Soil
4	Lambach Rail Tunnel	Rail	NATM	912	12.00	46.0	Rock
...
102	Seattle's Beacon Hill Light Rail	Rail	TBM	25000	6.40	7.0	Soil
103	Dulles airport, Washington	Walk	NATM	366	11.00	4.5	Soil
104	Hollywood section, LA Metro extension, Los Ang...	Metro	TBM	14000	6.70	25.0	Soil
105	Severomuisk railway tunnel, Baikal-Amur Line,	Rail	TBM	15343	4.56	1000.0	Rock
106	Haivan Pass Tunnel, Lang Co	Road	D&B	6280	11.90	890.0	Rock

107 rows x 7 columns

Εικόνα 1. Εισαγωγή dataset TFD_21 στο μοντέλο TFM1.

Όπως φαίνεται και στην Εικόνα 1, τα χαρακτηριστικά “tunnel use”, “cutting technique” και “rock/soil”, αποτελούν περιγραφικές μεταβλητές. Προκειμένου να εφαρμοστεί ο αλγόριθμος K-Means (βλ. Κεφ. 3.3.3 και 4.2.4) γίνεται μετασχηματισμός των

περιγραφικών τιμών σε αριθμητικές (Εικόνα 2), σύμφωνα με τις τιμές που ορίστηκαν και περιλαμβάνονται στον Πίνακα 4 (βλ. Κεφ. 4.2.4).

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Thirre Tunnel	7	6	5500	12.60	650.0	1
1	Lane cove Tunnel	2	6	130	8.00	17.0	1
2	Wienerwald tunnel, Vienna	4	4	13350	10.68	190.0	1
3	Lainzer railroad tunnel, Vienna	4	6	12800	5.55	26.0	2
4	Lambach Rail Tunnel	4	6	912	12.00	46.0	1
...
102	Seattle's Beacon Hill Light Rail	4	10	25000	6.40	7.0	2
103	Dulles airport, Washington	8	6	366	11.00	4.5	2
104	Hollywood section, LA Metro extension, Los Ang...	6	10	14000	6.70	25.0	2
105	Severomuisk railway tunnel, Baikal-Amur Line,	4	10	15343	4.56	1000.0	1
106	Haivan Pass Tunnel, Lang Co	7	3	6280	11.90	890.0	1

107 rows × 7 columns

Εικόνα 2. Μετασχηματισμός περιγραφικών τιμών σε αριθμητικές.

Για την αποφυγή της εκ παραδρομής απόδοσης μεγαλύτερης ή μικρότερης βαρύτητας σε συγκεκριμένο χαρακτηριστικό (βλ. Κεφ. 4.2.5), γίνεται κανονικοποίηση των τιμών, με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ των τιμών 0 και 1 (Εικόνα 3).

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Thirre Tunnel	0.857143	0.555556	0.045515	0.334477	0.259408	0.0
1	Lane cove Tunnel	0.142857	0.555556	0.000750	0.176672	0.006005	0.0
2	Wienerwald tunnel, Vienna	0.428571	0.333333	0.110954	0.268611	0.075260	0.0
3	Lainzer railroad tunnel, Vienna	0.428571	0.555556	0.106369	0.092624	0.009608	1.0
4	Lambach Rail Tunnel	0.428571	0.555556	0.007269	0.313894	0.017614	0.0
...
102	Seattle's Beacon Hill Light Rail	0.428571	1.000000	0.208069	0.121784	0.002002	1.0
103	Dulles airport, Washington	1.000000	0.555556	0.002718	0.279588	0.001001	1.0
104	Hollywood section, LA Metro extension, Los Ang...	0.714286	1.000000	0.116372	0.132075	0.009207	1.0
105	Severomuisk railway tunnel, Baikal-Amur Line,	0.428571	1.000000	0.127568	0.058662	0.399520	0.0
106	Haivan Pass Tunnel, Lang Co	0.857143	0.222222	0.052017	0.310463	0.355484	0.0

107 rows × 7 columns

Εικόνα 3. Κανονικοποίηση των τιμών με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ των τιμών 0 και 1.

Στη συνέχεια εφαρμόζεται ο αλγόριθμος K-Means στο dataset TFD_21, για τον οποίο ορίστηκε ως πλήθος συστάδων (clusters) $K = 4$. Στην Εικόνα 4, στη στήλη Cluster

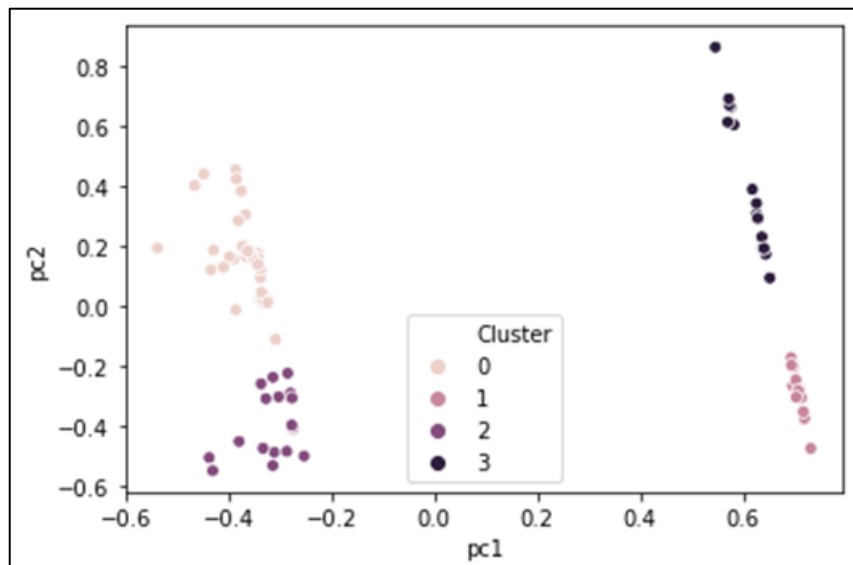
φαίνεται η συστάδα στην οποία εντάσσεται το κάθε περιστατικό. Οι συστάδες που διαμορφώνονται είναι οι 0, 1, 2 και 3.

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil	Cluster
0	Thirre Tunnel	0.857143	0.555556	0.045515	0.334477	0.259408	0.0	0
1	Lane cove Tunnel	0.142857	0.555556	0.000750	0.176672	0.006005	0.0	0
2	Wienerwald tunnel, Vienna	0.428571	0.333333	0.110954	0.268611	0.075260	0.0	0
3	Lainzer railroad tunnel, Vienna	0.428571	0.555556	0.106369	0.092624	0.009608	1.0	3
4	Lambach Rail Tunnel	0.428571	0.555556	0.007269	0.313894	0.017614	0.0	0
...
102	Seattle's Beacon Hill Light Rail	0.428571	1.000000	0.208069	0.121784	0.002002	1.0	3
103	Dulles airport, Washington	1.000000	0.555556	0.002718	0.279588	0.001001	1.0	1
104	Hollywood section, LA Metro extension, Los Ang...	0.714286	1.000000	0.116372	0.132075	0.009207	1.0	3
105	Severomuisk railway tunnel, Baikal-Amur Line,	0.428571	1.000000	0.127568	0.058662	0.399520	0.0	2
106	Haivan Pass Tunnel, Lang Co	0.857143	0.222222	0.052017	0.310463	0.355484	0.0	0

107 rows x 8 columns

Εικόνα 4. Εφαρμογή αλγορίθμου K-Means στο dataset TFD_21.

Προκειμένου να οπτικοποιηθούν τα αποτελέσματα της συσταδοποίησης, πραγματοποιείται μείωση της διαστατικότητας του dataset TFD_21, μέσω εφαρμογής της μεθόδου PCA (βλ. Κεφ. 4.4) και στη συνέχεια παράγεται το Διάγραμμα 10, στο οποίο απεικονίζονται τα στοιχεία του dataset, κατά τους άξονες pc1 και pc2, που αφορούν στις δύο κύριες συνιστώσες της PCA.



Διάγραμμα 10. Μείωση διαστατικότητας μέσω εφαρμογής PCA και απεικόνιση των δεδομένων υπό τις δύο κύριες συνιστώσες (pc1 και pc2).

Στη συνέχεια πραγματοποιείται η *σήμανση των συστάδων (cluster labeling)*. Για τη σήμανση υπολογίζεται αρχικά η συχνότητα εμφάνισης της κάθε συστάδας στο dataset, δηλαδή ο λόγος των περιστατικών που αντιστοιχούν σε κάθε συστάδα προς το σύνολο των περιστατικών. Δεδομένου ότι το dataset είναι αντιπροσωπευτικό δείγμα (βλ. Κεφ. 4.3) του πλήθους των περιστατικών αστοχίας που περιλαμβάνονται στη βάση δεδομένων, θεωρείται ότι η συχνότητα εμφάνισης του κάθε cluster, δηλώνει και την τάση αστοχίας των σηράγγων που παρουσιάζουν το ίδιο πρότυπο χαρακτηριστικών. Επομένως, κατατάσσοντας κατά φθίνουσα σειρά τη *σχετική συχνότητα αστοχίας (relative failure frequency)* του κάθε cluster (Εικόνα 5), προκύπτουν κατά αντιστοιχία τα labels “*Very High Frequency*”, “*High Frequency*”, “*Medium Frequency*” και “*Low Frequency*” (Εικόνα 6) και μία προς εξέταση σήραγγα μπορεί να ενταχθεί ομοίως στις κατηγορίες κινδύνου “*Very High Risk*”, “*High Risk*”, “*Medium Risk*” και “*Low Risk*”.

```
0    0.439252
2    0.214953
3    0.196262
1    0.149533
Name: Cluster, dtype: float64
```

Εικόνα 5. Φθίνουσα κατάταξη σχετικής συχνότητας αστοχίας των cluster (TFM1).

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil	Cluster	Cluster_Label	Relative_Failure_Frequency_%
0	Thirre Tunnel	Road	NATM	5500	12.60	650.0	Rock	0	Very High Frequency	43.925234
1	Lane cove Tunnel	Ventilation	NATM	130	8.00	17.0	Rock	0	Very High Frequency	43.925234
2	Wienerwald tunnel, Vienna	Rail	NATM/D&B	13350	10.68	190.0	Rock	0	Very High Frequency	43.925234
3	Lainzer railroad tunnel, Vienna	Rail	NATM	12800	5.55	26.0	Soil	3	Medium Frequency	19.626168
4	Lambach Rail Tunnel	Rail	NATM	912	12.00	46.0	Rock	0	Very High Frequency	43.925234
...
102	Seattle's Beacon Hill Light Rail	Rail	TBM	25000	6.40	7.0	Soil	3	Medium Frequency	19.626168
103	Dulles airport, Washington	Walk	NATM	366	11.00	4.5	Soil	1	Low Frequency	14.953271
104	Hollywood section, LA Metro extension, Los Ang...	Metro	TBM	14000	6.70	25.0	Soil	3	Medium Frequency	19.626168
105	Severomuisk railway tunnel, Baikal-Amur Line,	Rail	TBM	15343	4.56	1000.0	Rock	2	High Frequency	21.495327
106	Haivan Pass Tunnel, Lang Co	Road	D&B	6280	11.90	890.0	Rock	0	Very High Frequency	43.925234

107 rows x 10 columns

Εικόνα 6. Εισαγωγή των στηλών Cluster_Label και Relative_Failure_Frequency_% στο dataset (TFM1).

Τα αποτελέσματα της συσταδοποίησης και ο σχολιασμός τους, αποτελούν αντικείμενο του Κεφαλαίου 6.1.1.

5.1.2 Εκτίμηση Κινδύνου Αστοχίας Σήραγγας (TFM1_B)

Αρχικά ζητείται από το χρήστη να εισαχθούν τα χαρακτηριστικά της υπό εξέταση σήραγγας (Εικόνα 7), τα οποία είναι τα παρακάτω:

- α. “Tunnel Use”, με πιθανές τιμές: “Sewer”, “Ventilation”, “Hydro”, “Rail”, “Train (lightrail)”, “Metro”, “Road” και “Walk”.
- β. “Cutting Technique”, με πιθανές τιμές: “Hand”, “C&C”, “D&B”, “NATM/D&B”, “NATM/OPEN”, “NATM”, “NATM/TBM”, “TBM/D&B”, “TBM/C&C” και “TBM”.
- γ. “Length”, οποιαδήποτε τιμή σε μέτρα (m).
- δ. “Diameter”, οποιαδήποτε τιμή σε μέτρα (m).
- ε. “Overburden”, οποιαδήποτε τιμή σε μέτρα (m).
- στ. “Rock/Soil”, με πιθανές τιμές “Rock” και “Soil”.

Η σήραγγα προστίθεται στο dataset (με τις μετασχηματισμένες τιμές) ως εγγραφή, προκειμένου να εκτελεστούν στη συνέχεια οι υπολογισμοί που πραγματοποιούνται από το μοντέλο (Εικόνα 8).

```
# INPUT TEST TUNNEL FEATURES
pn = input('Give test tunnel Project Name: ')
tu = input('Give test tunnel Tunnel Use. 1 for Sewer, 2 for Ventilation, 3 for Hydro, 4 for Rail, 5 for Train (lightrail), 6 for Metro, 7 for Road and 8 for Walk: ')
ct = input('Give test tunnel Cutting Technique. 1 for HAND, 2 for C&C, 3 for D&B, 4 for NATM/D&B,5 for NATM/OPEN, 6 for NATM, 7 for NATM/TBM, 8 for TBM/D&B, 9 for TBM/C&C and 10 for TBM: ')
l = input('Give test tunnel Length in meters: ')
d = input('Give test tunnel Diameter in meters: ')
o = input('Give test tunnel Overburden in meters: ')
rs = input('Give test tunnel Ground. 1 for Rock and 2 for Soil: ')

Give test tunnel Project Name: Test Tunnel Case
Give test tunnel Tunnel Use. 1 for Sewer, 2 for Ventilation, 3 for Hydro, 4 for Rail, 5 for Train (lightrail), 6 for Metro, 7 for Road and 8 for Walk: 7
Give test tunnel Cutting Technique. 1 for HAND, 2 for C&C, 3 for D&B, 4 for NATM/D&B,5 for NATM/OPEN, 6 for NATM, 7 for NATM/TBM, 8 for TBM/D&B, 9 for TBM/C&C and 10 for TBM: 3
Give test tunnel Length in meters: 6000
Give test tunnel Diameter in meters: 11
Give test tunnel Overburden in meters: 800
Give test tunnel Ground. 1 for Rock and 2 for Soil: 1
```

Εικόνα 7. Εισαγωγή από το χρήστη των χαρακτηριστικών της υπό εξέταση σήραγγας (TFM1).

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Thirre Tunnel	0.857143	0.555556	0.045515	0.334477	0.259408	0.0
1	Lane cove Tunnel	0.142857	0.555556	0.000750	0.176672	0.006005	0.0
2	Wienerwald tunnel, Vienna	0.428571	0.333333	0.110954	0.268611	0.075260	0.0
3	Lainzer railroad tunnel, Vienna	0.428571	0.555556	0.106369	0.092624	0.009608	1.0
4	Lambach Rail Tunnel	0.428571	0.555556	0.007269	0.313894	0.017614	0.0
...
103	Dulles airport, Washington	1.000000	0.555556	0.002718	0.279588	0.001001	1.0
104	Hollywood section, LA Metro extension, Los Ang...	0.714286	1.000000	0.116372	0.132075	0.009207	1.0
105	Severomuisk railway tunnel, Baikal-Amur Line,	0.428571	1.000000	0.127568	0.058662	0.399520	0.0
106	Haivan Pass Tunnel, Lang Co	0.857143	0.222222	0.052017	0.310463	0.355484	0.0
107	Test Tunnel Case	0.857143	0.222222	0.049683	0.279588	0.319456	0.0

108 rows x 7 columns

Εικόνα 8. Προσθήκη της υπό εξέταση σήραγγας (Test Tunnel Case) ως εγγραφή (γραμμή 107) στο dataset (TFM1).

Στη συνέχεια, ομοίως με το TFM1_A, επειδή θα ακολουθήσει υπολογισμός ευκλείδειων αποστάσεων, προς αποφυγή εκ παραδρομής απόδοσης μεγαλύτερης ή μικρότερης βαρύτητας σε συγκεκριμένο χαρακτηριστικό (βλ. Κεφ. 4.2.5), γίνεται κανονικοποίηση των τιμών με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ των τιμών 0 και 1 (Εικόνα 9). Για την κανονικοποίηση λαμβάνονται υπόψη και οι τιμές που αφορούν στα χαρακτηριστικά της υπό εξέταση σήραγγας.

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Thirre Tunnel	0.857143	0.555556	0.045515	0.334477	0.259408	0.0
1	Lane cove Tunnel	0.142857	0.555556	0.000750	0.176672	0.006005	0.0
2	Wienerwald tunnel, Vienna	0.428571	0.333333	0.110954	0.268611	0.075260	0.0
3	Lainzer railroad tunnel, Vienna	0.428571	0.555556	0.106369	0.092624	0.009608	1.0
4	Lambach Rail Tunnel	0.428571	0.555556	0.007269	0.313894	0.017614	0.0
...
103	Dulles airport, Washington	1.000000	0.555556	0.002718	0.279588	0.001001	1.0
104	Hollywood section, LA Metro extension, Los Ang...	0.714286	1.000000	0.116372	0.132075	0.009207	1.0
105	Severomuisk railway tunnel, Baikal-Amur Line,	0.428571	1.000000	0.127568	0.058662	0.399520	0.0
106	Haivan Pass Tunnel, Lang Co	0.857143	0.222222	0.052017	0.310463	0.355484	0.0
107	Test Tunnel Case	0.857143	0.222222	0.049683	0.279588	0.319456	0.0

108 rows x 7 columns

Εικόνα 9. Κανονικοποίηση των τιμών, συμπεριλαμβανομένων και αυτών που αφορούν στην υπό εξέταση σήραγγα, με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ των τιμών 0 και 1.

Η ένταξη της υπό εξέταση σήραγγας σε μία από τις κατηγορίες κινδύνου αστοχίας, πραγματοποιείται με κριτήριο την ευκλείδεια απόσταση που παρουσιάζει το σημείο που την αντιπροσωπεύει στον εξα-διάστατο χώρο (έξι χαρακτηριστικά), από τα κεντροειδή των

συστάδων. Η ευκλείδεια απόσταση των δύο σημείων, δηλώνει την ομοιότητα του περιστατικού που εξετάζουμε με το υποθετικό περιστατικό που αντιστοιχεί στο κεντροειδές, το οποίο εκφράζει απολύτως το πρότυπο αστοχίας της συστάδας.

Η πιθανότητα αστοχίας υπολογίζεται με βάση την τιμή της ευκλείδειας απόστασης του σημείου από το πλησιέστερο κεντροειδές (Εικόνα 10) και τη σχετική συχνότητα αστοχίας (*relative failure frequency*) του αντίστοιχου cluster. Δεδομένου ότι το εύρος των κανονικοποιημένων τιμών, για όλα τα χαρακτηριστικά, είναι από 0 έως 1, η μέγιστη ευκλείδεια απόσταση που δύναται να παρουσιάσουν δύο εγγραφές – σημεία, είναι $(1+1+1+1+1)^{0,5} = 6^{0,5}$ και η ελάχιστη 0.

Επομένως, η εξεταζόμενη σήραγγα παρουσιάζει μεγαλύτερη ομοιότητα με την συστάδα από την οποία απέχει την ελάχιστη ευκλείδεια απόσταση και αν είναι πλησίον της συστάδας με label “Very High Frequency” / “High Frequency” / “Medium Frequency” / “Low Frequency” θα εμπίπτει αντιστοίχως στην κατηγορία κινδύνου “Very High Risk” / “High Risk” / “Medium Risk” / “Low Risk”.

Κατόπιν των ανωτέρω, είναι δυνατή η αναγωγή της ευκλείδειας απόστασης, σε ποσοστιαία ομοιότητα με το πλησιέστερο κεντροειδές ή με το κεντροειδές οποιασδήποτε συστάδας (*similarity to cluster’s centroid %*), μέσω της Εξίσωσης 16 και ο υπολογισμός της σχετικής πιθανότητας αστοχίας (*relative failure probability %*) μέσω της Εξίσωσης 17 (Εικόνα 11). Επίσης, η μέση τιμή της απόστασης του σημείου από τα κεντροειδή, η οποία δίνεται από την Εξίσωση 18, μπορεί να αποδώσει την ομοιότητα της προς εξέταση σήραγγας με αυτές που περιλαμβάνονται συνολικά στο dataset.

Με τον ίδιο τρόπο μπορεί να αποδοθεί και η ομοιότητα της υπό εξέταση σήραγγας με οποιαδήποτε από αυτές που περιλαμβάνονται στο dataset, σύμφωνα με την Εξίσωση 19 (Εικόνα 12). Επομένως, δύναται να εξαχθεί το περιστατικό αστοχίας σήραγγας το οποίο ομοιάζει στο μεγαλύτερο βαθμό με αυτό που εξετάζεται, δηλαδή τον “κοντινότερο γείτονα” (*nearest neighbor*).

$$\text{Similarity to Cluster's Centroid (\%)} = 100 - 100 * \text{Distance from Centroid} / 6^{0,5}$$

Εξίσωση 16

$$\text{Relative Failure Probability (\%)} =$$

$$\text{Cluster's Relative Failure Frequency (\%)} * \text{Similarity to Cluster's Centroid (\%)}$$

Εξίσωση 17

$$\text{Similarity to Dataset (\%)} = 100 - 100 * \Sigma \{ \text{Distance from each Centroid} / 6^{0.5} \} / 4$$

Εξίσωση 18

$$\text{Similarity to Case (\%)} = 100 - 100 * \text{Distance from Case} / 6^{0.5}$$

Εξίσωση 19

```
# CALCULATE DISTANCE OF THE TEST TUNNEL FROM EACH CENTROID.
distance_from_centroid_0 = distance.euclidean(test_tunnel_array, centroids[0])
distance_from_centroid_1 = distance.euclidean(test_tunnel_array, centroids[1])
distance_from_centroid_2 = distance.euclidean(test_tunnel_array, centroids[2])
distance_from_centroid_3 = distance.euclidean(test_tunnel_array, centroids[3])
distance_from_centroids = [distance_from_centroid_0, distance_from_centroid_1, distance_from_centroid_2, distance_from_centroid_3]
distance_from_centroids
```

Εικόνα 10. Υπολογισμός ευκλείδειας απόστασης υπό εξέταση σήραγγας από τα κεντροειδή των συστάδων (TFMI).

	Reshaped_Distance_From_Centroid	Cluster_Label	Similarity_to_Cluster's_Centroid_%
0	0.154527	Very High Frequency	84.5473
1	0.433493	Low Frequency	56.6507
2	0.346344	High Frequency	65.3656
3	0.559213	Medium Frequency	44.0787

Εικόνα 11. Αναγωγή (μετασχηματισμένης) απόστασης από το κεντροειδές σε ποσοστιαία ομοιότητα με το κεντροειδές (Similarity to Cluster's Centroid %) (TFMI).

chnique	Length	Diameter	Overburden	Rock_Soil	Cluster	Cluster_Label	Relative_Failure_Frequency_%	Reshaped_Distance	Test_Tunnel_Similarity_to_Case_%
NATM	5500	12.60	650.0	Rock	1	Very High Frequency	43.925234	0.140087	85.991253
NATM	130	8.00	17.0	Rock	1	Very High Frequency	43.925234	0.349417	65.058318
T/M/D&B	13350	10.68	190.0	Rock	1	Very High Frequency	43.925234	0.207976	79.202360
NATM	12800	5.55	26.0	Soil	2	Medium Frequency	19.626168	0.488016	51.198377
NATM	912	12.00	46.0	Rock	1	Very High Frequency	43.925234	0.254581	74.541908

Εικόνα 12. Υπολογισμός ομοιότητας της υπό εξέταση σήραγγας με κάθε περιστατικό του dataset και εισαγωγή της αντίστοιχης στήλης (Test Tunnel Similarity to Case %) (TFMI).

Στο παρόν σημείο θα πρέπει να επισημανθεί ότι στην εκτίμηση του κινδύνου αστοχίας υπό εξέταση σήραγγας, δεν λαμβάνεται υπόψη η συνέπεια της αστοχίας (οικονομική ή χρονική) καθότι οι αντίστοιχες στήλες της βάσης δεδομένων είναι πληρωμένες μόλις κατά 8% και 6% αντίστοιχα. Σε περίπτωση όμως που για τη εγγραφή που προκύπτει ως “κοντινότερος γείτονας” υπάρχουν αυτά τα στοιχεία, υπάρχει δυνατότητα υπολογισμού του κινδύνου ως γινόμενο της σχετικής πιθανότητας αστοχίας της υπό εξέταση σήραγγας επί την συνέπεια (οικονομική ή χρονική) της αστοχίας που κατεγράφη για τον “κοντινότερο γείτονα”.

5.2 Μοντέλο TFM2

Το μοντέλο TFM2 αποτελείται επίσης από δύο τμήματα. Το πρώτο τμήμα ενσωματώνει τον αλγόριθμο K-Prototypes, μέσω του οποίου διαμορφώνονται οι συστάδες (clusters) περιστατικών αστοχίας σηράγγων. Με βάση τα αποτελέσματα της συσταδοποίησης, γίνεται η συσχέτιση των συστάδων με τις ίδιες κατηγορίες κινδύνου που ορίστηκαν για το μοντέλο TFM1 και η σήμανση (labeling) των συστάδων με τα ίδια labels. Στο δεύτερο τμήμα ο χρήστης έχει τη δυνατότητα να εισάγει τα χαρακτηριστικά μίας υπό εξέταση σήραγγας και να εξάγει, αντίστοιχα με το TFM1, αποτελέσματα που αφορούν στην ομοιότητά της με τα περιστατικά που περιλαμβάνονται στο TFD_21 και στον κίνδυνο αστοχίας της. Συγκεκριμένα, εξάγεται η πλησιέστερη περίπτωση αστοχίας σήραγγας (*κοντινότερος γείτονας – nearest neighbor*), ο βαθμός ομοιότητάς της με αυτή, με το cluster στο οποίο υπάγεται και με το σύνολο των περιστατικών του TFD_21, αλλά και η σχετική πιθανότητα αστοχίας.

5.2.1 Συσταδοποίηση (TFM2_A)

Αρχικά εισάγεται το προς ανάλυση dataset TFD_21 στο μοντέλο (Εικόνα 13).

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Thirre Tunnel	Road	NATM	5500	12.60	650.0	Rock
1	Lane cove Tunnel	Ventilation	NATM	130	8.00	17.0	Rock
2	Wienerwald tunnel, Vienna	Rail	NATM/D&B	13350	10.68	190.0	Rock
3	Lainzer railroad tunnel, Vienna	Rail	NATM	12800	5.55	26.0	Soil
4	Lambach Rail Tunnel	Rail	NATM	912	12.00	46.0	Rock
...
102	Seattle's Beacon Hill Light Rail	Rail	TBM	25000	6.40	7.0	Soil
103	Dulles airport, Washington	Walk	NATM	366	11.00	4.5	Soil
104	Hollywood section, LA Metro extension, Los Ang...	Metro	TBM	14000	6.70	25.0	Soil
105	Severomuisk railway tunnel, Baikal-Amur Line,	Rail	TBM	15343	4.56	1000.0	Rock
106	Haivan Pass Tunnel, Lang Co	Road	D&B	6280	11.90	890.0	Rock

107 rows x 7 columns

Εικόνα 13. Εισαγωγή dataset TFD_21 στο μοντέλο TFM2.

Όπως φαίνεται στην Εικόνα 13, τα χαρακτηριστικά “tunnel use”, “cutting technique” και “rock/soil”, αποτελούν περιγραφικές μεταβλητές. Για την εφαρμογή του αλγορίθμου K-Prototypes δεν απαιτείται ο μετασχηματισμός των περιγραφικών τιμών των σε αριθμητικές,

καθότι για αυτές γίνεται υπολογισμός του μέτρου ανομοιότητας (βλ. Κεφ. 3.3.4). Για την αποφυγή της εκ παραδρομής απόδοσης μεγαλύτερης ή μικρότερης βαρύτητας σε συγκεκριμένο αριθμητικό χαρακτηριστικό (βλ. Κεφ. 4.2.5), γίνεται κανονικοποίηση των αριθμητικών τιμών, με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ των τιμών 0 και 1 (Εικόνα 14).

	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Road	NATM	0.045515	0.334477	0.259408	Rock
1	Ventilation	NATM	0.000750	0.176672	0.006005	Rock
2	Rail	NATM/D&B	0.110954	0.268611	0.075260	Rock
3	Rail	NATM	0.106369	0.092624	0.009608	Soil
4	Rail	NATM	0.007269	0.313894	0.017614	Rock
...
102	Rail	TBM	0.208069	0.121784	0.002002	Soil
103	Walk	NATM	0.002718	0.279588	0.001001	Soil
104	Metro	TBM	0.116372	0.132075	0.009207	Soil
105	Rail	TBM	0.127568	0.058662	0.399520	Rock
106	Road	D&B	0.052017	0.310463	0.355484	Rock

107 rows x 6 columns

Εικόνα 14. Κανονικοποίηση των αριθμητικών τιμών με τη μέθοδο ελάχιστου–μέγιστου, μεταξύ 0 και 1.

Ακολούθως, εφαρμόζεται ο αλγόριθμος K-Prototypes στο dataset TFD_21, για τον οποίο ορίστηκε ως πλήθος συστάδων (clusters) $K = 4$. Στην Εικόνα 15, στη στήλη *Cluster* φαίνεται η συστάδα στην οποία εντάσσεται το κάθε περιστατικό. Οι συστάδες που διαμορφώνονται είναι οι 0, 1, 2 και 3.

	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil	Cluster
0	Road	NATM	0.045515	0.334477	0.259408	Rock	1
1	Ventilation	NATM	0.000750	0.176672	0.006005	Rock	1
2	Rail	NATM/D&B	0.110954	0.268611	0.075260	Rock	1
3	Rail	NATM	0.106369	0.092624	0.009608	Soil	3
4	Rail	NATM	0.007269	0.313894	0.017614	Rock	1
...
102	Rail	TBM	0.208069	0.121784	0.002002	Soil	3
103	Walk	NATM	0.002718	0.279588	0.001001	Soil	1
104	Metro	TBM	0.116372	0.132075	0.009207	Soil	3
105	Rail	TBM	0.127568	0.058662	0.399520	Rock	0
106	Road	D&B	0.052017	0.310463	0.355484	Rock	1

107 rows x 7 columns

Εικόνα 15. Εφαρμογή αλγορίθμου K-Prototypes στο dataset TFD_21.

Στη συνέχεια πραγματοποιείται ο υπολογισμός της σχετικής συχνότητας αστοχίας των συστάδων και η σήμανση των συστάδων (cluster labeling). Η μέθοδος υπολογισμού είναι η ίδια με εκείνη που εφαρμόζεται στο TFM1 και αναφέρεται στο Κεφάλαιο 5.1.1. (Εικόνες 16 και 17).

```

1    0.457944
3    0.271028
0    0.224299
2    0.046729
Name: Cluster, dtype: float64

```

Εικόνα 16. Φθίνουσα κατάταξη σχετικής συχνότητας αστοχίας των clusters (TFM2).

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil	Cluster	Cluster_Label	Relative_Failure_Frequency_%
0	Thirre Tunnel	Road	NATM	5500	12.60	650.0	Rock	1	Very High Frequency	45.794393
1	Lane cove Tunnel	Ventilation	NATM	130	8.00	17.0	Rock	1	Very High Frequency	45.794393
2	Wienerwald tunnel, Vienna	Rail	NATM/D&B	13350	10.68	190.0	Rock	1	Very High Frequency	45.794393
3	Lainzer railroad tunnel, Vienna	Rail	NATM	12800	5.55	26.0	Soil	3	High Frequency	27.102804
4	Lambach Rail Tunnel	Rail	NATM	912	12.00	46.0	Rock	1	Very High Frequency	45.794393
...
102	Seattle's Beacon Hill Light Rail	Rail	TBM	25000	6.40	7.0	Soil	3	High Frequency	27.102804
103	Dulles airport, Washington	Walk	NATM	366	11.00	4.5	Soil	1	Very High Frequency	45.794393
104	Hollywood section, LA Metro extension, Los Ang...	Metro	TBM	14000	6.70	25.0	Soil	3	High Frequency	27.102804
105	Severomuisk railway tunnel, Baikal-Amur Line,	Rail	TBM	15343	4.56	1000.0	Rock	0	Medium Frequency	22.429907
106	Haivan Pass Tunnel, Lang Co	Road	D&B	6280	11.90	890.0	Rock	1	Very High Frequency	45.794393

107 rows x 10 columns

Εικόνα 17. Εισαγωγή των στηλών Cluster Label και Relative Failure Frequency % στο dataset (TFM2).

Τα αποτελέσματα της συσταδοποίησης και ο σχολιασμός τους, αποτελούν αντικείμενο του Κεφαλαίου 6.2.1.

5.2.2 Εκτίμηση Κινδύνου Αστοχίας Σήραγγας (TFM2_B)

Αρχικά εισάγονται από το χρήστη τα χαρακτηριστικά της υπό εξέταση σήραγγας (Εικόνα 18). Τα χαρακτηριστικά και οι αντίστοιχες πιθανές τιμές που δύναται να εισαχθούν είναι τα ίδια με του TFM1_A και αναφέρονται στο Κεφ. 5.1.2.. Η σήραγγα προστίθεται στο dataset ως εγγραφή, προκειμένου να εκτελεστούν στη συνέχεια οι υπολογισμοί που πραγματοποιούνται από το μοντέλο (Εικόνα 19).

```
# INPUT TEST TUNNEL FEATURES
pn = input('Give test tunnel Project Name: ')
tu = input('Give test tunnel Tunnel Use. 1 for Sewer, 2 for Ventilation, 3 for Hydro, 4 for Rail, 5 for Train (lightrail), 6 for
ct = input('Give test tunnel Cutting Technique. 1 for HAND, 2 for C&C, 3 for D&B, 4 for NATM/D&B,5 for NATM/OPEN, 6 for NATM, 7
l = input('Give test tunnel Length in meters: ')
d = input('Give test tunnel Diameter in meters: ')
o = input('Give test tunnel Overburden in meters: ')
rs = input('Give test tunnel Ground. 1 for Rock and 2 for Soil: ')

Give test tunnel Project Name: Test Tunnel Case
Give test tunnel Tunnel Use. 1 for Sewer, 2 for Ventilation, 3 for Hydro, 4 for Rail, 5 for Train (lightrail), 6 for Metro, 7 f
or Road and 8 for Walk: 7
Give test tunnel Cutting Technique. 1 for HAND, 2 for C&C, 3 for D&B, 4 for NATM/D&B,5 for NATM/OPEN, 6 for NATM, 7 for NATM/TB
M, 8 for TBM/D&B, 9 for TBM/C&C and 10 for TBM: 3
Give test tunnel Length in meters: 6000
Give test tunnel Diameter in meters: 11
Give test tunnel Overburden in meters: 800
Give test tunnel Ground. 1 for Rock and 2 for Soil: 1
```

Εικόνα 18. Εισαγωγή από το χρήστη των χαρακτηριστικών της υπό εξέταση σήραγγας (TFM2).

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Thirre Tunnel	Road	NATM	5500	12.6	650	Rock
1	Lane cove Tunnel	Ventilation	NATM	130	8	17	Rock
2	Wienerwald tunnel, Vienna	Rail	NATM/D&B	13350	10.68	190	Rock
3	Lainzer railroad tunnel, Vienna	Rail	NATM	12800	5.55	26	Soil
4	Lambach Rail Tunnel	Rail	NATM	912	12	46	Rock
...
103	Dulles airport, Washington	Walk	NATM	366	11	4.5	Soil
104	Hollywood section, LA Metro extension, Los Ang...	Metro	TBM	14000	6.7	25	Soil
105	Severomuisk railway tunnel, Baikal-Amur Line,	Rail	TBM	15343	4.56	1000	Rock
106	Haivan Pass Tunnel, Lang Co	Road	D&B	6280	11.9	890	Rock
107	Test Tunnel Case	Road	D&B	6000	11	800	Rock

108 rows x 7 columns

Εικόνα 19. Προσθήκη της υπό εξέταση σήραγγας (Test Tunnel Case) ως εγγραφή (γραμμή 107) dataset (TFM2).

Στη συνέχεια, ομοίως με το TFM2_A, επειδή θα ακολουθήσει υπολογισμός ευκλείδειων αποστάσεων, προς αποφυγή εκ παραδρομής απόδοσης μεγαλύτερης ή μικρότερης βαρύτητας σε συγκεκριμένο αριθμητικό χαρακτηριστικό (βλ. Κεφ. 4.2.5), γίνεται κανονικοποίηση των αριθμητικών τιμών με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ

των τιμών 0 και 1. Για την κανονικοποίηση λαμβάνονται υπόψη και οι τιμές που αφορούν στα χαρακτηριστικά της υπό εξέταση σήραγγας (Εικόνα 20).

	Project	Tunnel_Use	Cutting_Technique	Length	Diameter	Overburden	Rock_Soil
0	Thirre Tunnel	Road	NATM	0.045515	0.334477	0.259408	Rock
1	Lane cove Tunnel	Ventilation	NATM	0.000750	0.176672	0.006005	Rock
2	Wienerwald tunnel, Vienna	Rail	NATM/D&B	0.110954	0.268611	0.075260	Rock
3	Lainzer railroad tunnel, Vienna	Rail	NATM	0.106369	0.092624	0.009608	Soil
4	Lambach Rail Tunnel	Rail	NATM	0.007269	0.313894	0.017614	Rock
...
103	Dulles airport, Washington	Walk	NATM	0.002718	0.279588	0.001001	Soil
104	Hollywood section, LA Metro extension, Los Ang...	Metro	TBM	0.116372	0.132075	0.009207	Soil
105	Severomuisk railway tunnel, Baikal-Amur Line,	Rail	TBM	0.127568	0.058662	0.399520	Rock
106	Haivan Pass Tunnel, Lang Co	Road	D&B	0.052017	0.310463	0.355484	Rock
107	Test Tunnel Case	Road	D&B	0.049683	0.279588	0.319456	Rock

108 rows × 7 columns

Εικόνα 20. Κανονικοποίηση των αριθμητικών τιμών, συμπεριλαμβανομένων και αυτών που αφορούν στην υπό εξέταση σήραγγα, με τη μέθοδο ελάχιστου – μέγιστου, μεταξύ των τιμών 0 και 1.

Για την ένταξη της υπό εξέταση σήραγγας σε μία από τις κατηγορίες κινδύνου αστοχίας και για τον υπολογισμό της πιθανότητας αστοχίας στο μοντέλο TFM2 ακολουθείτε διαφορετική μέθοδος από εκείνη του TFM1, καθότι ο αλγόριθμος K-Prototypes υπολογίζει κατά τη διαμόρφωση των συστάδων, την “απόσταση” των σημείων, ως το άθροισμα δύο όρων. Ο πρώτος όρος αφορά στην τετραγωνική ευκλείδεια απόσταση και ο δεύτερος όρος στο μέτρο ανομοιότητας. Επομένως, κατ’ αντιστοιχία με τον αλγόριθμο, στο μοντέλο ορίζεται ως “απόσταση” της υπό εξέταση σήραγγας από το κεντροειδές, το άθροισμα της ευκλείδειας απόστασης των σημείων στον τρισ-διάστατο χώρο (τρία αριθμητικά γνωρίσματα) και του μέτρου ανομοιότητας αυτών. Δεδομένου όμως, ότι το εύρος των κανονικοποιημένων τιμών, για όλα τα αριθμητικά γνωρίσματα, είναι από 0 έως 1, η μέγιστη ευκλείδεια απόσταση που δύναται να παρουσιάσουν δύο εγγραφές – σημεία, είναι $(1+1+1)^{0.5} = 3^{0.5}$ και η ελάχιστη 0, ενώ για τα περιγραφικά γνωρίσματα, το μέγιστο μέτρο ανομοιότητας που μπορούν να παρουσιάσουν δύο εγγραφές είναι $1+1+1=3$ και το ελάχιστο 0 (Εικόνα 21).

Κατόπιν των ανωτέρω, είναι δυνατή η αναγωγή της “απόστασης”, σε ποσοστιαία ομοιότητα με το πλησιέστερο κεντροειδές ή με το κεντροειδές της κάθε συστάδας (*similarity to cluster’s centroid %*), μέσω της Εξίσωσης 20 και ο υπολογισμός της σχετικής πιθανότητας αστοχίας μέσω της Εξίσωσης 17 (Εικόνα 22). Τέλος, ομοίως με το TFM1, προκύπτει η ομοιότητα της υπό εξέταση σήραγγας με αυτές που περιλαμβάνονται συνολικά

στο dataset μέσω της Εξίσωσης 21, με οποιαδήποτε από αυτές που περιλαμβάνονται στο dataset μέσω της Εξίσωσης 22 και μέσω φθίνουσας κατάταξης αυτών προκύπτει και ο “κοντινότερος γείτονας” (Εικόνα 23).

$$\text{Similarity to Cluster's Centroid (\%)} = 100 - 100 * \text{Distance from Centroid} / (3^{0,5} + 3)$$

Εξίσωση 20

$$\text{Similarity to Dataset (\%)} = 100 - 100 * \Sigma \{ \text{Distance from each Centroid} / (3^{0,5} + 3) \} / 4$$

Εξίσωση 21

$$\text{Similarity to Case (\%)} = 100 - 100 * \text{Distance from Case} / (3^{0,5} + 3)$$

Εξίσωση 22

```
df['Distance_from_Centroid_0'] = df_num['Distance_from_Centroid_0'] + df_desc['Distance_from_Centroid_0']
df['Distance_from_Centroid_1'] = df_num['Distance_from_Centroid_1'] + df_desc['Distance_from_Centroid_1']
df['Distance_from_Centroid_2'] = df_num['Distance_from_Centroid_2'] + df_desc['Distance_from_Centroid_2']
df['Distance_from_Centroid_3'] = df_num['Distance_from_Centroid_3'] + df_desc['Distance_from_Centroid_3']
df
```

Εικόνα 21. Υπολογισμός “απόστασης” υπό εξέταση σήραγγας από τα κεντροειδή των συστάδων.

	Reshaped_Distance_From_Centroid	Cluster_Label	Similarity_to_Cluster's_Centroid_%
0	0.263013	Very High Frequency	73.6987
1	0.550396	Low Frequency	44.9604
2	0.476797	Medium Frequency	52.3203
3	0.704509	High Frequency	29.5491

Εικόνα 22. Αναγωγή “απόστασης” από το κεντροειδές σε ποσοστιαία ομοιότητα με το κεντροειδές (Similarity to Cluster's Centroid %).

```
Test_Tunnel_Similarity_to_Case_%
0      77.146041
1      50.686859
2      52.409559
3      28.861745
4      51.253032
..
102    28.397108
103    29.800006
104    29.207336
105    52.503525
106    98.996090
[107 rows x 13 columns]
```

Εικόνα 23. Υπολογισμός ομοιότητας της υπό εξέταση σήραγγας με κάθε περιστατικό του dataset και εισαγωγή της αντίστοιχης στήλης (Test Tunnel Similarity to Case %) (TFM2).

6 Αποτελέσματα Ανάλυσης Δεδομένων

Στο παρόν κεφάλαιο παρουσιάζονται τα αποτελέσματα εφαρμογής των μοντέλων στο dataset. Στόχος είναι η ποιοτική αλλά και η ποσοτική αξιολόγηση των αποτελεσμάτων της ανάλυσης, με σκοπό την εντοπισμό των προτύπων (ή μοτίβων) χαρακτηριστικών που οδηγούν σε μικρότερο ή μεγαλύτερο κίνδυνο αστοχίας. Επισημαίνεται, ότι η δυνατότητα εκτίμησης προτύπων, υφίσταται λόγω της εφαρμογής των μοντέλων μηχανικής μάθησης και συγκεκριμένα των αλγορίθμων συσταδοποίησης. Μία αντίστοιχη διαδικασία στατιστικής επεξεργασίας των στοιχείων, χωρίς τη χρήση αλγορίθμων, θα ήταν ιδιαίτερα χρονοβόρα και απαιτητική και πιθανώς θα οδηγούσε σε εσφαλμένα συμπεράσματα, καθότι θα ήταν εξαιρετικά δύσκολο να αναγνωριστούν οι συσχετίσεις των χαρακτηριστικών του παρόντος πολυδιάστατου (έξι χαρακτηριστικά γνωρίσματα = έξι διαστάσεις) προβλήματος.

6.1 Αποτελέσματα Μοντέλου TFM1

6.1.1 Αποτελέσματα Συσταδοποίησης (TFM1_A)

Από την εφαρμογή του μοντέλου TFM1 (τμήμα TFM1_A), το οποίο χρησιμοποιεί τον αλγόριθμο K-Means, προέκυψε η διαμόρφωση των τεσσάρων (4) συστάδων (clusters) περιστατικών αστοχίας σηράγγων. Η ομοιότητα μίας προς εξέταση σήραγγας με το cluster συγκεκριμένης σχετικής συχνότητας αστοχίας, υποδηλώνει αντίστοιχα την κατηγορία κινδύνου στην οποία εμπίπτει. Επομένως, προκύπτουν αντιστοίχως με τα clusters τέσσερις (4) κατηγορίες κινδύνου. Οι συστάδες οι οποίες προέκυψαν από την εφαρμογή στο dataset, το οποίο περιλαμβάνει συνολικά 118 περιστατικά, φαίνονται στον Πίνακα 5. Με βάση τον αριθμό των περιστατικών, προέκυψαν οι σχετικές συχνότητες αστοχίας και τα αντίστοιχα labels των συστάδων (βλ. Κεφ. 5.1.1.).

<i>Cluster Label</i>	<i>Number of Incidents</i>	<i>Relative Failure Frequency</i>
Low Frequency	16	15,0 %
Medium Frequency	21	19,6 %
High Frequency	23	21,5 %
Very High Frequency	47	43,9 %

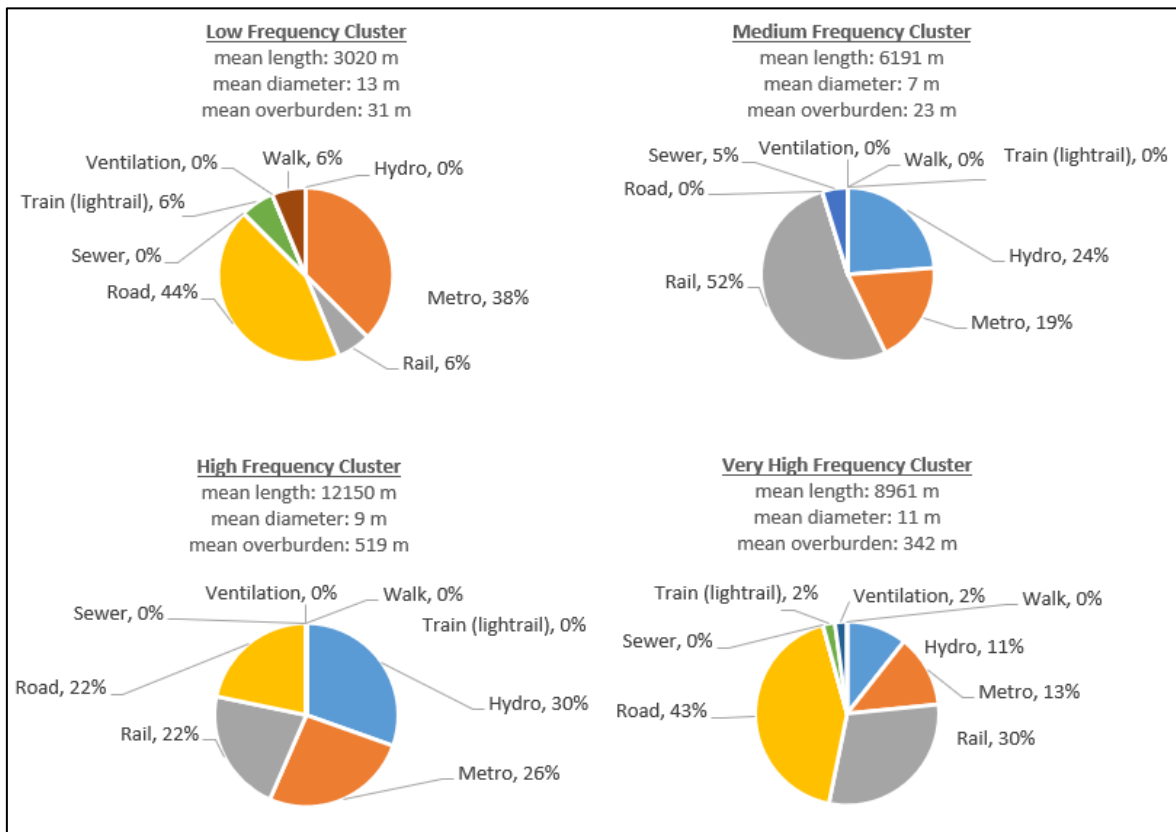
Πίνακας 5. Σχετικές συχνότητες αστοχίας και τα αντίστοιχα labels των clusters που προέκυψαν από τη εφαρμογή του TFM1 στο dataset.

Μία πρώτη παρατήρηση που μπορεί να γίνει άμεσα είναι ότι το cluster “Very High Frequency” παρουσιάζει σχεδόν διπλάσια συχνότητα εμφάνισης στο dataset σε σχέση με τα υπόλοιπα clusters, τα clusters “High Frequency” και “Medium Frequency” σχεδόν την ίδια και την χαμηλότερη, αλλά όχι με μεγάλη διαφορά από τα υπόλοιπα, παρουσιάζει το cluster “Low Frequency”. Από αυτή την παρατήρηση εξάγεται το συμπέρασμα ότι οι σήραγγες με το πρότυπο χαρακτηριστικών του cluster “Very High Frequency” είναι αυτές που ουσιαστικά παρουσιάζουν υψηλό κίνδυνο αστοχίας, ενώ ο κίνδυνος αστοχίας για σήραγγες που εμπίπτουν στα υπόλοιπα clusters είναι όμοιος.

Στον Πίνακα 6, φαίνεται η ποσοστιαία εμφάνιση των τιμών που αφορούν στα περιγραφικά χαρακτηριστικά των σηράγγων ανά cluster, τα οποία στη συνέχεια απεικονίζονται και διαγραμματικά.

<i>Χαρακτηριστικό</i>	<i>Τιμή</i>	<i>Cluster</i>			
		<i>Low Frequency</i>	<i>Medium Frequency</i>	<i>High Frequency</i>	<i>Very High Frequency</i>
Tunnel Use	Hydro	0 %	24 %	30 %	11 %
	Metro	38 %	19 %	26 %	13 %
	Rail	6 %	52 %	22 %	30 %
	Road	44 %	0 %	22 %	43 %
	Sewer	0 %	5 %	0 %	0 %
	Train (light rail)	6 %	0 %	0 %	2 %
	Ventilation	0 %	0 %	0 %	2 %
	Walk	6 %	0 %	0 %	0 %
Cutting Technique	C&C	25 %	0 %	0 %	2 %
	D&B	13 %	0 %	0 %	23 %
	HAND	6 %	0 %	0 %	0 %
	NATM	50 %	5 %	0 %	68 %
	NATM/D&B	0 %	0 %	0 %	4 %
	NATM/OPEN	6 %	0 %	0 %	0 %
	NATM/TBM	0 %	0 %	0 %	2 %
	TBM	0 %	95 %	100 %	0 %
Rock/Soil	Rock	0 %	0 %	100 %	100 %
	Soil	100 %	100 %	0 %	0 %

Πίνακας 6. Ποσοστιαία αναλογία εμφάνισης τιμών που αφορούν στα περιγραφικά χαρακτηριστικά των σηράγγων, ανά cluster (TFM1).



Εικόνα 24. Ποσοστιαία εμφάνιση τιμών του περιγραφικού χαρακτηριστικού tunnel use ανά cluster (TFMI).

Στα διαγράμματα της Εικόνας 24 φαίνεται η ποσοστιαία εμφάνιση των τιμών του περιγραφικού χαρακτηριστικού “tunnel use”. Το χαρακτηριστικό αυτό, είναι γνωστό εκ των προτέρων ότι συνδέεται με το αριθμητικό χαρακτηριστικό “diameter” με υψηλό βαθμό συσχέτισης, καθότι το είδος της σήραγγας είναι ο καθοριστικός παράγοντας που λαμβάνεται υπόψη για το σχεδιασμό της σήραγγας και ως εκ τούτου για τον ορισμό της απαιτούμενης διατομής. Αυτό φαίνεται και στα διαγράμματα, όπου παρατηρείται ότι στα clusters που απαντώνται χαμηλές τιμές διαμέτρου σήραγγας (μέση τιμή 9 και 7 m για τα clusters “High Frequency” και “Medium Frequency”) απαντώνται και τα υψηλότερα ποσοστά τιμών Hydro και Sewer. Αντιθέτως, στα clusters που απαντώνται μεγάλες τιμές διαμέτρου (μέση τιμή 11 και 13 m για τα clusters “Low Frequency” και “Very High Frequency”) απαντάται το μεγαλύτερο ποσοστό εμφάνιση της τιμής “Road”. Το χαρακτηριστικό “tunnel use”, συμπεριλήφθη στην ανάλυση προκειμένου να αξιολογηθεί αν το είδος της σήραγγας είναι ένα χαρακτηριστικό το οποίο συσχετίζεται με την ποιότητα σχεδίασης ή/και κατασκευής

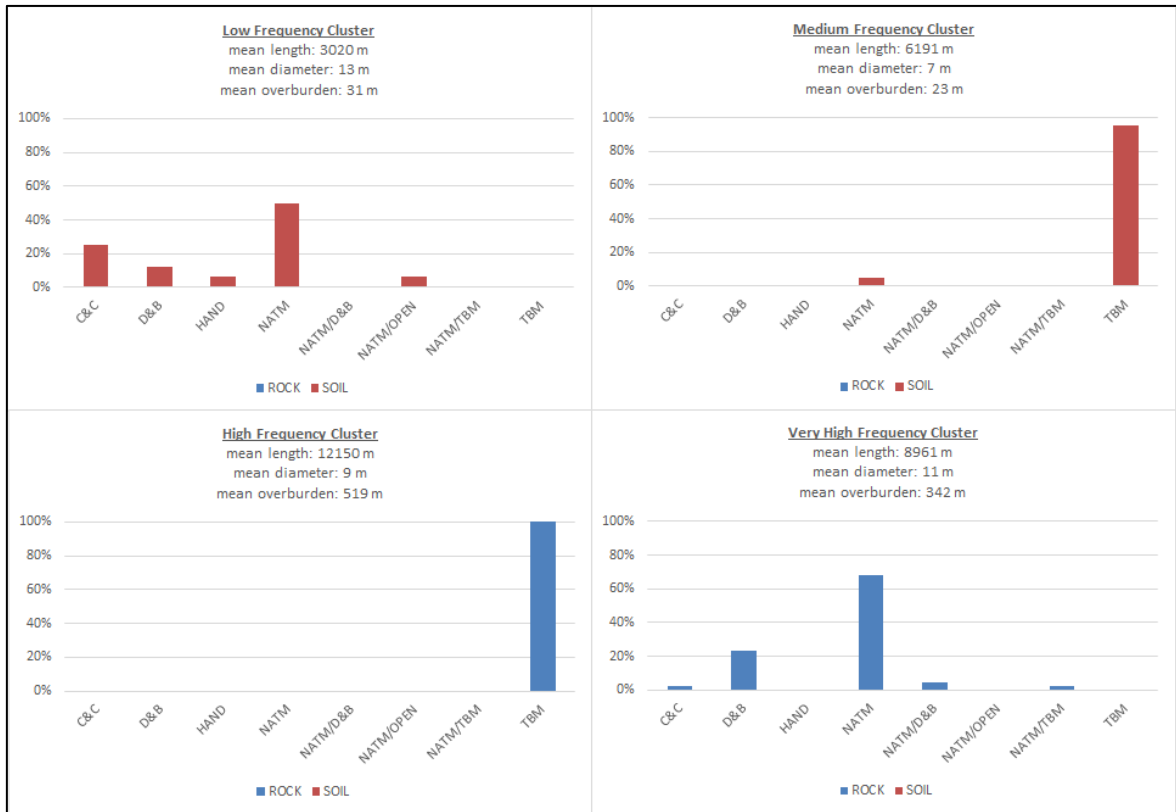
μίας σήραγγας. Από την παρούσα ανάλυση όμως δεν προκύπτει κάποιο αποτέλεσμα που να υποδεικνύει οποιαδήποτε τέτοια συσχέτιση.

Όσον αφορά στο μήκος της σήραγγας, παρατηρείται ότι το μέσο μήκος των σηράγγων που περιλαμβάνονται στα cluster “High Frequency” και “Very High Frequency”, είναι μεγαλύτερο σε σχέση με εκείνο που παρατηρείται στα cluster “Low Frequency” και “Medium Frequency”. Επομένως, ο ισχυρισμός ότι όσο μεγαλύτερο είναι το μήκος της σήραγγας τόσο μεγαλύτερη είναι η πιθανότητα εμφάνισης αστοχίας, επιβεβαιώνεται και από την ανάλυση. Το ίδιο ακριβώς επιβεβαιώνεται και για το χαρακτηριστικό “overburden”, το οποίο δηλώνει και το μέγεθος της τάσης που ασκούν τα υπερκείμενα εδάφη, όπου παρατηρούμε ότι στις κατηγορίες υψηλής συχνότητας αστοχίας απαντώνται κατά πολύ μεγαλύτερες τιμές πάχους υπερκείμενων.

Ιδιαίτερο όμως ενδιαφέρον παρουσιάζει η διαγραμματική απεικόνιση των χαρακτηριστικών “cutting technique” και “rock/soil”, τα οποία όπως αποδείχτηκε και από την PCA (βλ. Κεφ. 4.4.2) μπορούν να θεωρηθούν τα σημαντικότερα χαρακτηριστικά που λαμβάνονται υπόψη στην ανάλυση. Στην **Εικόνα 25** φαίνονται τα διαγράμματα που αφορούν σε αυτά τα χαρακτηριστικά και στη συσχέτιση αυτών. Από τη συσταδοποίηση προκύπτει πως μεγαλύτερη συχνότητα αστοχίας (“Very High Frequency” και “High Frequency”) παρουσιάζουν οι σήραγγες οι οποίες διανοίγονται σε βράχο (τιμή: “rock”). Συγκεκριμένα, αυτές που διανοίγονται με τις συμβατικές μεθόδους διάνοιξης (τιμές: “NATM”, “D&B”) σε βράχο υπάγονται στο cluster “Very High Frequency” και αυτές που διανοίγονται με TBM σε βράχο υπάγονται στο cluster “High Frequency”. Μικρότερη συχνότητα αστοχίας παρουσιάζουν οι σήραγγες οι οποίες διανοίγονται σε έδαφος (τιμή: “soil”) και συγκεκριμένα, αυτές που διανοίγονται με τις συμβατικές μεθόδους διάνοιξης (τιμές: “NATM”, “D&B”, “C&C”) υπάγονται στο cluster “Low Frequency” και αυτές που διανοίγονται με TBM υπάγονται στο cluster “Medium Frequency”.

Συνοψίζοντας, στον **Πίνακα 7** παρουσιάζονται τα πρότυπα (μοτίβα) χαρακτηριστικών, μη λαμβάνοντας υπόψη το χαρακτηριστικό “tunnel use” το οποίο θεωρείται ότι καλύπτεται από το χαρακτηριστικό “diameter”, που προκύπτουν για κάθε cluster και ως εκ τούτου, που δύναται να οδηγήσουν στην αντίστοιχη κατηγορία κινδύνου. Επισημαίνεται ότι η χρήση του **Πίνακα 7** για την εξέταση σήραγγας, ενδείκνυται μόνο για σήραγγες οι οποίες συμφωνούν πλήρως με το πρότυπο χαρακτηριστικών κάποιου cluster.

Αν η συμφωνία δεν ικανοποιείται και για τα πέντε χαρακτηριστικά, η εξέταση θα πρέπει να γίνεται με την εφαρμογή του μοντέλου.



Εικόνα 25. Διαγραμματική απεικόνιση χαρακτηριστικών cutting technique και rock/soil, ανά cluster (TFM1).

Χαρακτηριστικό	Cluster			
	Low Frequency	Medium Frequency	High Frequency	Very High Frequency
Cutting Technique	Conventional Methods	TBM	TBM	Conventional Methods
Length	< 7 km	< 7km	> 7km	> 7km
Diameter	> 10 m	< 10 m	< 10 m	> 10 m
Overburden	< 40 m	< 40 m	> 40 m	> 40 m
Rock/Soil	Soil	Soil	Rock	Rock

Πίνακας 7. Πρότυπα (μοτίβα) χαρακτηριστικών που προκύπτουν από την διαμόρφωση των cluster (TFM1).

6.1.2 Αποτελέσματα Εκτίμησης Κινδύνου Αστοχίας Σήραγγας (TFM1_B)

Για την αξιολόγηση των αποτελεσμάτων που προκύπτουν από την εφαρμογή του τμήματος TFM1_B του μοντέλου TFM1, πραγματοποιήθηκε εκτίμηση κινδύνου αστοχίας για μία σήραγγα, η οποία είναι ήδη κατασκευασμένη και γνωρίζουμε ότι παρουσίασε αστοχία κατά την κατασκευή. Πρόκειται για τη σιδηροδρομική (τιμή: “rail”) σήραγγα Ampurdan στην Ισπανία, η οποία κατά τη διάρκεια κατασκευής της, εμφάνισε σημαντικές αστοχίες λόγω υπερφόρτισης της προσωρινής υποστήριξης (OSF) [32]. Η συγκεκριμένη σήραγγα δεν ακολουθεί επακριβώς κάποιο από τα πρότυπα αστοχίας που παρατηρήθηκαν στο Κεφ. 6.1.1. και παρουσιάζονται συνοπτικά στον Πίνακα 7. Για το λόγο αυτό θεωρήθηκε κατάλληλη για την αξιολόγηση της λειτουργίας εκτίμησης κινδύνου αστοχίας που πραγματοποιείται από την εφαρμογή του μοντέλου εισάγοντας τα χαρακτηριστικά της.

Στην Εικόνα 26 φαίνονται τα χαρακτηριστικά της σήραγγας Ampurdan (τιμές: “rail”, “NATM”, “length” 449 m, “diameter” 12 m, “overburden” 33 m, “rock”). Με βάση τα χαρακτηριστικά της και σύμφωνα με τα πρότυπα του Πίνακα 7, η σήραγγα θα μπορούσε να ενταχθεί στην κατηγορία κινδύνου “High Risk”, καθότι συμφωνεί με το πρότυπο του cluster “High Frequency” σε 3 από τα 5 χαρακτηριστικά. Όπως όμως επισημάνθηκε η χρήση του πίνακα ενδείκνυται μόνο όταν υπάρχει συμφωνία 5/5 με κάποιο πρότυπο. Επομένως, θα εξεταστεί και μέσω εφαρμογής του TFM1_B. Στην Εικόνα 27 φαίνονται τα αποτελέσματα της εφαρμογής.

```
Give test tunnel Project Name: Ampurdan Tunnel Spain
Give test tunnel Tunnel Use. 1 for Sewer, 2 for Ventilation, 3 for Hydro, 4 for Rail, 5 for Train (lightrail), 6 for Metro, 7 for Road and 8 for Walk: 4
Give test tunnel Cutting Technique. 1 for HAND, 2 for C&C, 3 for D&B, 4 for NATM/D&B, 5 for NATM/OPEN, 6 for NATM, 7 for NATM/TBM, 8 for TBM/D&B, 9 for TBM/C&C and 10 for TBM: 6
Give test tunnel Length in meters: 649
Give test tunnel Diameter in meters: 12
Give test tunnel Overburden in meters: 33
Give test tunnel Ground. 1 for Rock and 2 for Soil: 1
```

Εικόνα 26. Εισαγωγή χαρακτηριστικών υπό εξέταση σήραγγας στο μοντέλο TFM1.

```
The test tunnel can fall to the Very High Risk Category and bears a similarity of 89.12013840421298 % to the Very High Frequency Cluster's Centroid and a similarity of 99.76945744203302 % to the Nearest Neighbor Tunnel Failure Case.
The similarity of the test tunnel to the whole dataset is 69.71680751643456 %
The Nearest Neighbor Tunnel Failure Case is the following :
Project Lambach Rail Tunnel
Tunnel_Use Rail
Cutting_Technique NATM
Length 912
Diameter 12
Overburden 46
Rock_Soil Rock
Cluster 1
Cluster_Label Very High Frequency
Relative_Failure_Frequency_% 43.9252
Reshaped_Distance 0.00230543
Test_Tunnel_Similarity_to_Case_% 99.7695
The Relative Failure Probabilty of the test tunnel case is 39.14622901867299 %.
```

Εικόνα 27. Αποτελέσματα εφαρμογής του μοντέλου TFM1 στην υπό εξέταση σήραγγα.

Σύμφωνα με τα ανωτέρω αποτελέσματα, η σήραγγα εμπίπτει στην κατηγορία κινδύνου “Very High Risk” και φέρει ομοιότητα με το κεντροειδές του cluster “Very High Frequency” 89,1%, ομοιότητα με τον “κοντινότερο γείτονα” 99,8% και σχετική πιθανότητα αστοχίας 39,1%. Επισημαίνεται, ότι η υψηλότερη σχετική πιθανότητα αστοχίας που δύναται να φέρει μία σήραγγα που εμπίπτει σε αυτή την κατηγορία κινδύνου είναι 43,9% (ισότιμη δηλαδή με την συχνότητα αστοχίας του cluster). Επομένως, λαμβάνοντας υπόψη την υψηλή ομοιότητα με το κεντροειδές του cluster, την υψηλή σχετική πιθανότητα αστοχίας αλλά κυρίως την υψηλή ομοιότητα με τον “κοντινότερο γείτονα”, η σήραγγα θα μπορούσε να χαρακτηριστεί από εξαιρετικά υψηλό κίνδυνο αστοχίας, το οποίο επιβεβαιώθηκε και στην πραγματικότητα.

6.2 Αποτελέσματα Μοντέλου TFM2

6.2.1 Αποτελέσματα Συσταδοποίησης (TFM2_A)

Από την εφαρμογή του μοντέλου TFM2 (τμήμα TFM2_A), το οποίο χρησιμοποιεί τον αλγόριθμο K-Prototypes, προέκυψε η διαμόρφωση των συστάδων (clusters) περιστατικών αστοχίας σηράγγων. Ο αριθμός των clusters είχε εξ’ αρχής οριστεί σε $K = 4$, όπως και για την εφαρμογή το μοντέλου TFM1 (βλ. Κεφ. 6.1.1), και μια υπό εξέταση σήραγγα δύναται να εμπίπτει σε μία από τις τέσσερις (4) αντίστοιχες κατηγορίες κινδύνου.

Οι συστάδες οι οποίες προέκυψαν από την εφαρμογή στο dataset, το οποίο περιλαμβάνει συνολικά 118 περιστατικά, φαίνονται στον Πίνακα 8. Με βάση τον αριθμό των περιστατικών προέκυψαν οι σχετικές συχνότητες αστοχίας και τα αντίστοιχα labels των συστάδων (βλ. Κεφ. 5.2.1.).

<i>Cluster Label</i>	<i>Number of Incidents</i>	<i>Relative Failure Frequency</i>
Low Frequency	5	4,7 %
Medium Frequency	24	22,4 %
High Frequency	29	27,1 %
Very High Frequency	49	45,8 %

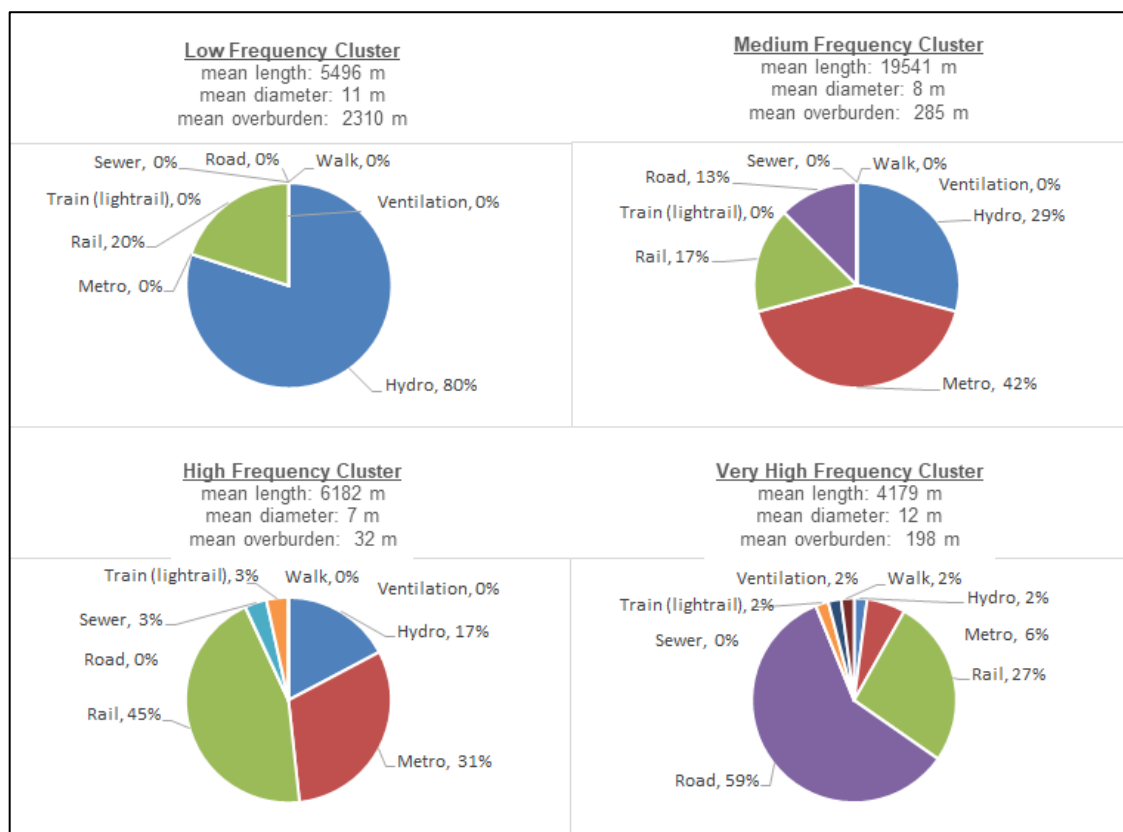
Πίνακας 8. Σχετικές συχνότητες αστοχίας και τα αντίστοιχα labels των clusters που προέκυψαν από τη εφαρμογή του TFM2 στο dataset.

Μία παρατήρηση που μπορεί να γίνει άμεσα είναι ότι το cluster “Very High Frequency” παρουσιάζει σχεδόν διπλάσια συχνότητα εμφάνισης, σε σχέση με τα clusters “High Frequency” και “Medium Frequency”, ενώ το cluster “Low Frequency” παρουσιάζει εξαιρετικά χαμηλότερη συχνότητα από τα υπόλοιπα. Από αυτή την παρατήρηση εξάγεται το συμπέρασμα ότι οι σήραγγες με το πρότυπο χαρακτηριστικών του cluster “Very High Frequency” είναι αυτές που ουσιαστικά παρουσιάζουν υψηλό κίνδυνο αστοχίας, ο κίνδυνος αστοχίας για σήραγγες που εμπίπτουν στα clusters “High Frequency” και “Medium Frequency” είναι όμοιος, ενώ για το cluster “Low Frequency” είναι πολύ χαμηλός.

Στον Πίνακα 9, φαίνεται η ποσοστιαία εμφάνιση των τιμών που αφορούν στα περιγραφικά χαρακτηριστικά των σηράγγων ανά cluster, τα οποία στη συνέχεια απεικονίζονται και διαγραμματικά.

Χαρακτηριστικό	Τιμή	Cluster			
		Low Frequency	Medium Frequency	High Frequency	Very High Frequency
Tunnel Use	Hydro	80 %	29 %	17 %	2 %
	Metro	0 %	42 %	31 %	6 %
	Rail	20 %	17 %	45 %	27 %
	Road	0 %	13 %	0 %	59 %
	Sewer	0 %	0 %	3 %	0 %
	Train (light rail)	0 %	0 %	3 %	2 %
	Ventilation	0 %	0 %	0 %	2 %
	Walk	0 %	0 %	0 %	2 %
Cutting Technique	C&C	0 %	0 %	7 %	6 %
	D&B	40 %	17 %	0 %	14 %
	HAND	0 %	0 %	3 %	0 %
	NATM	0 %	8 %	17 %	69 %
	NATM/D&B	0 %	0 %	0 %	4 %
	NATM/OPEN	0 %	0 %	0 %	2 %
	NATM/TBM	0 %	4 %	0 %	0 %
	TBM	60 %	71 %	72 %	4 %
Rock/Soil	Rock	100 %	100 %	3 %	82 %
	Soil	0 %	0 %	97 %	8 %

Πίνακας 9. Ποσοστιαία αναλογία εμφάνισης τιμών που αφορούν στα περιγραφικά χαρακτηριστικά των σηράγγων, ανά cluster (TFM1).



Εικόνα 28. Ποσοστιαία εμφάνιση τιμών του περιγραφικού χαρακτηριστικού tunnel use ανά cluster (TFM2).

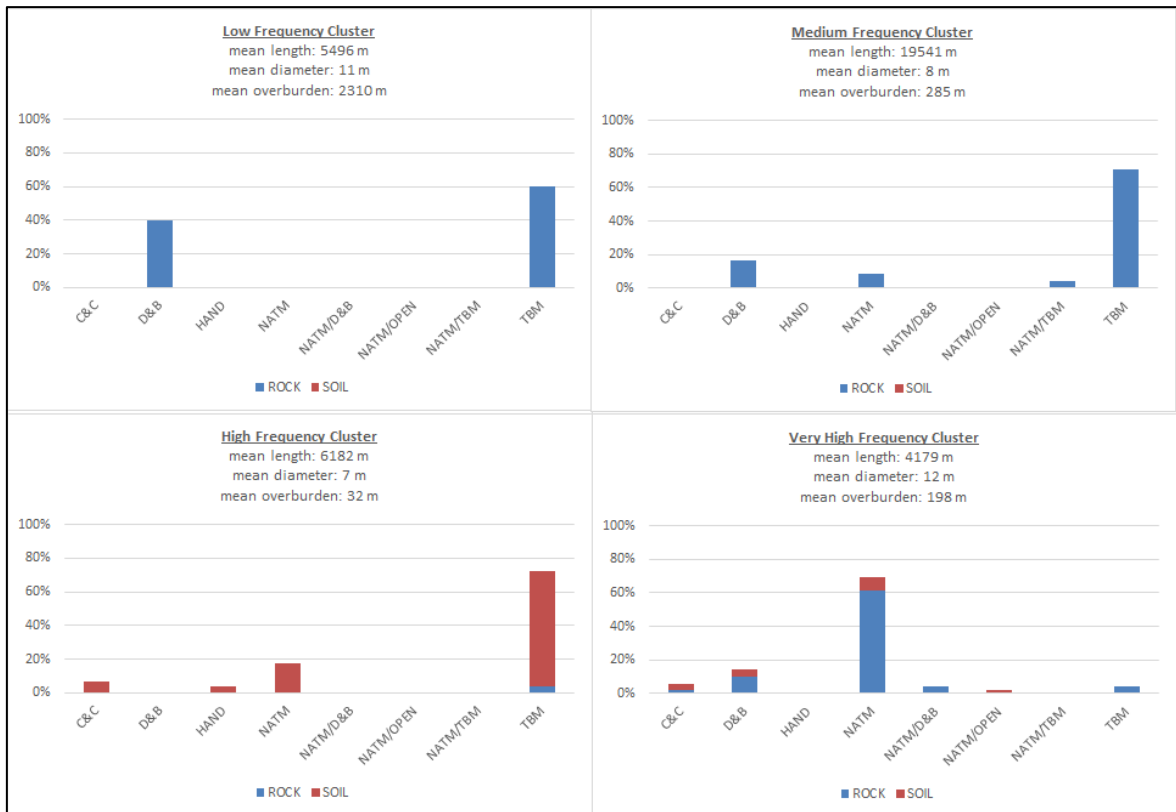
Στα διαγράμματα της Εικόνας 28 φαίνεται η ποσοστιαία εμφάνιση των τιμών του περιγραφικού χαρακτηριστικού “tunnel use” και η μέση τιμή των αριθμητικών χαρακτηριστικών “length”, “diameter” και “overburden”. Παρατηρούμε ότι το “Low Frequency” cluster έχει μέση τιμή “overburden” 2310m, βασισμένο σε 5 περιστατικά αστοχίας τα οποία είχαν τις υψηλότερες τιμές “overburden” του dataset (1550 - 2500m). Επομένως η διαμόρφωση του cluster θεωρούμε ότι αποδίδεται κατά κύριο λόγο σε αυτό το χαρακτηριστικό. Το χαρακτηριστικό “tunnel use”, όπως προαναφέρθηκε στο Κεφ. 6.1.1 συμπεριλήφθη στην ανάλυση προκειμένου να αξιολογηθεί αν η ποιότητα του σχεδιασμού ή/και της κατασκευής μίας σήραγγας σχετίζεται με τον προορισμό χρήσης της. Από την παρούσα ανάλυση όμως δεν προκύπτει κάποιο αποτέλεσμα που να υποδεικνύει οποιαδήποτε τέτοια συσχέτιση.

Όσον αφορά στο μήκος της σήραγγας, παρατηρούμε ότι το μέσο μήκος των σηράγγων που περιλαμβάνονται στο cluster “Medium Frequency”, είναι κατά πολύ μεγαλύτερο σε σχέση με εκείνο που παρατηρείται σε όλα τα υπόλοιπα cluster. Επομένως, ο

ισχυρισμός ότι όσο μεγαλύτερο είναι το μήκος της σήραγγας τόσο μεγαλύτερη είναι η πιθανότητα εμφάνισης αστοχίας, δεν επιβεβαιώνεται και από την ανάλυση. Το ίδιο ακριβώς δεν επιβεβαιώνεται ούτε για το χαρακτηριστικό “overburden”, το οποίο δηλώνει και το μέγεθος της τάσης που ασκούν τα υπερκείμενα εδάφη.

Ιδιαίτερο όμως ενδιαφέρον παρουσιάζει η διαγραμματική απεικόνιση των χαρακτηριστικών “cutting technique” και “rock/soil”, τα οποία όπως αποδείχτηκε και από την PCA (βλ. Κεφ. 4.4.2) μπορούν να θεωρηθούν τα σημαντικότερα χαρακτηριστικά που λαμβάνονται υπόψη στην ανάλυση. Στην Εικόνα 29 φαίνονται τα διαγράμματα που αφορούν σε αυτά τα χαρακτηριστικά και στη συσχέτιση αυτών. Από τη συσταδοποίηση προκύπτει πως μεγαλύτερη συχνότητα αστοχίας (cluster: “Very High Frequency”) παρουσιάζουν οι σήραγγες οι οποίες διανοίγονται σε βράχο (τιμή: rock) με συμβατικές μεθόδους διάνοιξης (τιμές: “NATM”, “D&B”). Αυτές που διανοίγονται, κατά κύριο λόγο, με TBM σε έδαφος (τιμή: “soil”) υπάγονται στο cluster “High Frequency”. Μικρότερη συχνότητα αστοχίας παρουσιάζουν οι σήραγγες οι οποίες διανοίγονται σε βράχο με TBM, οι οποίες υπάγονται στο cluster “Medium Frequency”. Όσον αφορά στις τιμές των χαρακτηριστικών “cutting technique” και “rock/soil” του cluster “Low Frequency,” δεν δύναται να εξαχθεί ασφαλές συμπέρασμα καθότι η διαμόρφωση του cluster προέκυψε ως αποτέλεσμα της ύπαρξης ακραίων τιμών για τα χαρακτηριστικό “overburden”.

Συνοψίζοντας, στον Πίνακα 10 παρουσιάζονται τα πρότυπα (μοτίβα) χαρακτηριστικών που προκύπτουν για κάθε cluster και ως εκ τούτου, που δύναται να οδηγήσουν στην αντίστοιχη κατηγορία κινδύνου. Επισημαίνεται ότι η χρήση του Πίνακα 10 για την εξέταση σήραγγας, ενδείκνυται μόνο για σήραγγες οι οποίες συμφωνούν πλήρως με το πρότυπο χαρακτηριστικών κάποιου cluster. Αν η συμφωνία δεν αφορά και στα πέντε χαρακτηριστικά, η εξέταση θα πρέπει να γίνεται με την εφαρμογή του μοντέλου.



Εικόνα 29. Διαγραμματική απεικόνιση χαρακτηριστικών cutting technique και rock/soil, ανά cluster (TFM2).

Χαρακτηριστικό	Cluster			
	Low Frequency	Medium Frequency	High Frequency	Very High Frequency
Cutting Technique	All methods	TBM	TBM	Conventional Methods
Length	< 2,5km	> 8km	< 8km	All lengths
Diameter	> 7m	< 10 m	< 10 m	> 10 m
Overburden	> 2000 m	> 40 m	< 40 m	All depths
Rock/Soil	Rock	Rock	Soil	Rock

Πίνακας 10. Πρότυπα (μοτίβα) χαρακτηριστικών που προκύπτουν από τη διαμόρφωση των cluster (TFM2).

6.2.2 Αποτελέσματα Εκτίμησης Κινδύνου Αστοχίας Σήραγγας (TFM1_B)

Για την αξιολόγηση των αποτελεσμάτων που προκύπτουν από την εφαρμογή του τμήματος TFM2_B του μοντέλου TFM2, πραγματοποιήθηκε εκτίμηση κινδύνου αστοχίας για την ίδια σήραγγα που εξετάστηκε κατά την αξιολόγηση του TFM1_B, ήτοι η

σιδηροδρομική (τιμή: “rail”) σήραγγα Ampurdan στην Ισπανία, η οποία κατά τη διάρκεια κατασκευής της, εμφάνισε σημαντικές αστοχίες λόγω υπερφόρτισης της προσωρινής υποστήριξης (τιμή: “OSF”) [32]. Η συγκεκριμένη σήραγγα δεν ακολουθεί επακριβώς κάποιο από τα πρότυπα αστοχίας που παρατηρήθηκαν στο **Κεφ. 6.2.1.** και παρουσιάζονται συνοπτικά στον Πίνακα 10. Για το λόγο αυτό θεωρήθηκε κατάλληλη για την αξιολόγηση της λειτουργίας εκτίμησης κινδύνου αστοχίας που πραγματοποιείται από την εφαρμογή του μοντέλου εισάγοντας τα χαρακτηριστικά της.

Στην **Εικόνα 30** φαίνονται τα χαρακτηριστικά της σήραγγας Ampurdan (τιμές χαρακτηριστικών: “rail”, “NATM”, “length” 449 m, “diameter” 12 m, “overburden” 33 m, “rock”). Με βάση τα χαρακτηριστικά της και σύμφωνα με τα πρότυπα του **Πίνακα 10**, η σήραγγα θα μπορούσε να ενταχθεί στην κατηγορία κινδύνου “High Risk”, καθώς συμφωνεί με το πρότυπο του cluster “High Frequency” σε 5 από τα 5 χαρακτηριστικά. Όπως επισημάνθηκε η χρήση του πίνακα ενδείκνυται μόνο όταν υπάρχει συμφωνία 5/5 με κάποιο πρότυπο. Για επαλήθευση, η σήραγγα θα εξεταστεί και μέσω εφαρμογής του TFM1_B. Στην **Εικόνα 31** φαίνονται τα αποτελέσματα της εφαρμογής.

```
Give test tunnel Project Name: Ampurdan Tunnel Spain
Give test tunnel Tunnel Use. 1 for Sewer, 2 for Ventilation, 3 for Hydro, 4 for Rail, 5 for Train (lightrail), 6 for Metro, 7 for Road and 8 for Walk: 4
Give test tunnel Cutting Technique. 1 for HAND, 2 for C&C, 3 for D&B, 4 for NATM/D&B, 5 for NATM/OPEN, 6 for NATM, 7 for NATM/TBM, 8 for TBM/D&B, 9 for TBM/C&C and 10 for TBM: 6
Give test tunnel Length in meters: 649
Give test tunnel Diameter in meters: 12
Give test tunnel Overburden in meters: 33
Give test tunnel Ground. 1 for Rock and 2 for Soil: 1
```

Εικόνα 30. Εισαγωγή χαρακτηριστικών υπό εξέταση σήραγγας στο μοντέλο TFM2.

```
The test tunnel can fall to the Very High Risk Category and bears a similarity of 77.35990948238512 % to the Very High Frequency Cluster's Centroid and a similarity of 99.86307847654268 % to the Nearest Neighbor Tunnel Failure Case.

The similarity of the test tunnel to the whole dataset is 61.041383840272275 %

The Nearest Neighbor Tunnel Failure Case is the following :

Project                Lambach Rail Tunnel
Tunnel_Use              Rail
Cutting_Technique       NATM
Length                  912
Diameter                12
Overburden              46
Rock_Soil               Rock
Cluster                 2
Cluster_Label           Very High Frequency
Relative_Failure_Frequency_% 44.8598
Reshaped_Distance       0.00136922
Test_Tunnel_Similarity_to_Case_% 99.8631

The Relative Failure Probability of the test tunnel case is 34.703510795836316 %.
```

Εικόνα 31. Αποτελέσματα εφαρμογής του μοντέλου TFM2 στην υπό εξέταση σήραγγα.

Σύμφωνα με τα ανωτέρω αποτελέσματα, η σήραγγα εμπίπτει στην κατηγορία κινδύνου “Very High Risk” και φέρει ομοιότητα με το κεντροειδές του cluster “Very High

Frequency” 89,1%, ομοιότητα με τον “κοντινότερο γείτονα” 99,8% και σχετική πιθανότητα αστοχίας 39,1%. Επισημαίνεται, ότι η υψηλότερη σχετική πιθανότητα αστοχίας που δύναται να φέρει μία σήραγγα που εμπίπτει σε αυτή την κατηγορία κινδύνου είναι 43,9% (ισότιμη δηλαδή με την συχνότητα αστοχίας του cluster). Επομένως, λαμβάνοντας υπόψη την υψηλή ομοιότητα με το κεντροειδές του cluster, την υψηλή σχετική πιθανότητα αστοχίας αλλά κυρίως την υψηλή ομοιότητα με τον κοντινότερο γείτονα, η σήραγγα θα μπορούσε να χαρακτηριστεί από εξαιρετικά υψηλό κίνδυνο αστοχίας, το οποίο επιβεβαιώθηκε και στην πραγματικότητα.

7 Πρόβλεψη Τύπου Αστοχίας Σήραγγας (TFM1_C και TFM2_C)

Σκοπός του παρόντος σκέλους της εργασίας είναι η εξέταση της δυνατότητας πρόβλεψης του τύπου αστοχίας υπό εξέταση σήραγγας, από τα μοντέλα TFM1 και TFM2. Τα μοντέλα, πραγματοποιούν ανάλυση του dataset TFD_21, το οποίο περιλαμβάνει 107 εγγραφές, με τιμές για συγκεκριμένα χαρακτηριστικά (βλ. Κεφ. 4.3). Ένα επιπλέον χαρακτηριστικό το οποίο μέχρι στιγμής δεν έχει αξιοποιηθεί, είναι το “failure type” (τύπος αστοχίας), για το οποίο στη βάση δεδομένων υπάρχουν τιμές για όλες τις εγγραφές που περιλαμβάνονται στο TFD_21. Οι τιμές που δύνανται να εμφανίζει το χαρακτηριστικό και η συντομογραφία τους φαίνονται στον Πίνακα 11. Επομένως, το χαρακτηριστικό συμπεριελήφθη στην ανάλυση, προκειμένου να εκπαιδευτούν τα μοντέλα μας και στη συνέχεια να πραγματοποιηθούν προβλέψεις, βάση των οποίων υπολογίστηκε και η ακρίβεια τους.

<i>Συντομογραφία</i>	<i>Failure Type</i>
FAC	Face Collapse
OSF	Overstressing Support Failure
FIR	Fire
CRW	Crown Wedge Failure
LS	Lateral Support Failure
OTH	Other

Πίνακας 11. Τιμές χαρακτηριστικού τύπος αστοχίας (failure type) και αντίστοιχες συντομογραφίες.

7.1 Μεθοδολογία Πρόβλεψης Τύπου Αστοχίας

Από την εφαρμογή των τμημάτων TFM1_A και TFM2_A των μοντέλων, προκύπτουν τα τέσσερα (4) clusters περιστατικών αστοχίας. Κάθε cluster περιλαμβάνει περιστατικά όπου προκλήθηκαν διάφοροι τύπους αστοχίας. Σε κάθε cluster, άλλοι τύποι αστοχίας εμφανίζονται περισσότερο και άλλοι λιγότερο, δηλαδή έχουν μία συγκεκριμένη *σχετική συχνότητα εμφάνισης (relative failure type frequency)*. Επιπλέον, από τα τμήματα

TFM1_B και TFM2_B των μοντέλων, προκύπτει ο “κοντινότερος γείτονας” μίας υπό εξέταση σήραγγας, ο οποίος έχει παρουσιάσει επίσης συγκεκριμένο τύπο αστοχίας.

Για την προσέγγιση που ακολουθείται θεωρεί ότι ο τύπος αστοχίας που δύναται να παρουσιάσει μια σήραγγα, σχετίζεται με το πρότυπο χαρακτηριστικών της. Επομένως, λαμβάνοντας υπόψη τον τύπο αστοχίας που παρουσίασαν οι σήραγγες με το ίδιο πρότυπο, δηλαδή η σήραγγες του ίδιου cluster, αλλά και η σήραγγα που παρουσιάζει την μεγαλύτερη ομοιότητα, δηλαδή ο “κοντινότερος γείτονας”, δύναται να υπολογιστεί η *σχετική πιθανότητα εμφάνισης ενός συγκεκριμένου τύπου αστοχίας (relative probability of a certain failure type)* για μία υπό εξέταση σήραγγα. Το τμήμα των μοντέλων που πραγματοποιεί τον υπόψη υπολογισμό είναι το τρίτο τμήμα (TFM1_C και TFM2_C) και η μεθοδολογία που χρησιμοποιείται, όπως και η κωδικοποίηση αυτής, είναι ίδια και για τα δύο μοντέλα.

Για τον υπολογισμό της σχετικής πιθανότητας εμφάνισης συγκεκριμένου τύπου αστοχίας πρώτα υπολογίζεται η σχετική συχνότητα εμφάνισης της κάθε τιμής του χαρακτηριστικού “failure type” στο cluster στο οποίο εμπίπτει η προς εξέταση σήραγγα (Εικόνα 32).

	Failure_Type	Failure_Type_Frequency
0	FAC	0.568182
1	OSF	0.340909
2	CRW	0.068182
3	FAC/OSF	0.022727

Εικόνα 32 Σχετική συχνότητα εμφάνισης τύπων αστοχίας (στήλη *Failure_Type_Frequency*) που εμφανίζονται στο cluster στο οποίο εμπίπτει μια προς εξέταση σήραγγα.

Επιπλέον, για την υπό εξέταση σήραγγα έχει προκύψει από την εφαρμογή του δεύτερου τμήματος του μοντέλου, ο “κοντινότερος γείτονας” (Nearest Neighbor ή N.N.) για τον οποίο είναι γνωστός επίσης ο τύπος αστοχίας που παρουσίασε, όπως φαίνεται στην Εικόνα 33.

```
The test tunnel can fall to the Very High Risk Category and bears a similarity of 87.24324113802626 % to the Very High Frequency Cluster's Centroid and a similarity of 94.5561060523031 % to the Nearest Neighbor Tunnel Failure Case.

The similarity of the test tunnel to the whole dataset is 67.38950356971691 %

The Nearest Neighbor Tunnel Failure Case is the following :

Project                Mingyazi tunnel
Tunnel_Use              Road
Cutting_Technique       NATM
Length                  4944
Diameter                 12.1
Overburden               320
Rock_Soil                Rock
Cluster                 2
Cluster_Label           Very High Frequency
Relative_Failure_Frequency_% 45.8333
Reshaped_Distance        0.0544389
Test_Tunnel_Similarity_to_Case_% 94.5561
Failure_Type             OSF
Name: 0, dtype: object

The Relative Failure Probability of the test tunnel case is 39.98648552159537 %.
```

Εικόνα 33. Στοιχεία "κοντινότερου γείτονα" της υπό εξέταση σήραγγας (τύπος αστοχίας OSF: Overstressing Support Failure).

Για την πρόβλεψη, θεωρείται πως η σχετική πιθανότητα εμφάνισης ενός συγκεκριμένου τύπου αστοχίας, είναι ανάλογη της σχετικής συχνότητας εμφάνισης του συγκεκριμένου τύπου αστοχίας, στο cluster στο οποίο εμπίπτει η σήραγγα (Εξίσωση 23). Δηλαδή, μία υπό εξέταση σήραγγα η οποία ακολουθεί σε ορισμένο βαθμό το πρότυπο ενός cluster, θεωρείται ότι δύναται να εμφανίσει όλους τους τύπους αστοχίας που εμφανίζονται στο cluster, αλλά με πιθανότητα ανάλογη της σχετικής συχνότητας εμφάνισης του συγκεκριμένου τύπου.

Relative Probability of a Certain Failure Type (%)

$$= \frac{\text{Relative Failure Type Frequency of Failure Type} * \text{Relative Failure Probability (\%)}}{\Sigma(\text{Relative Failure Type Frequency of All Possible Types})}$$

Εξίσωση 23

Όμως, δεδομένου ότι κατά την διαδικασία συσταδοποίησης των περιστατικών, στους υπολογισμούς που εκτελούνται από τους αλγόριθμους δεν συμμετέχει το χαρακτηριστικό "failure type", οι συστάδες διαμορφώνονται βάσει των συσχετίσεων που παρατηρούνται στα υπόλοιπα χαρακτηριστικά και το εν λόγω χαρακτηριστικό ουσιαστικά λαμβάνεται υπόψη κατά τη σήμανση των συστάδων, με μία όμως μόνο πιθανή τιμή, την τιμή "failure" που εκφράζει την εμφάνιση αστοχίας γενικά. Επομένως, το γεγονός ότι μία υπό εξέταση σήραγγα εμπίπτει σε ένα cluster, δηλώνει ότι σε αυτήν αναγνωρίζεται το πρότυπο

χαρακτηριστικών που οδηγεί σε αστοχία γενικά και όχι σε εμφάνιση συγκεκριμένου τύπου αστοχίας. Αντιθέτως, ο βαθμός εγγύτητας μίας υπό εξέταση σήραγγας στον “κοντινότερο γείτονα”, δηλώνει τον βαθμό ταύτισης του προτύπου χαρακτηριστικών της, με αυτό του “κοντινότερου γείτονα”, ο οποίος εμφάνισε συγκεκριμένο τύπο αστοχίας. Επομένως, θεωρείται ότι η πιθανότητα εμφάνισης του συγκεκριμένου τύπου αστοχίας που παρουσιάζει ο “κοντινότερος γείτονας” (Nearest Neighbor ή N.N.), είναι αυξημένη σε σχέση με αυτήν που προκύπτει αναλογικά της συχνότητας εμφάνισης της στο cluster που εμπίπτει η εξεταζόμενη σήραγγα και η θεώρηση αυτή εκφράζεται με την ύπαρξη ενός συντελεστή βαρύτητας W , ο οποίος δύναται να λάβει τιμές μεγαλύτερες της μονάδας. Ο ορισμός της τιμής του συντελεστή βαρύτητας W που εν τέλει όμως χρησιμοποιήθηκε για την τελική διαμόρφωση των μοντέλων, είναι αντικείμενο του **Κεφαλαίου 7.2**. Λαμβάνοντας υπόψη τον συντελεστή W στην **Εξίσωση 23**, για την περίπτωση του τύπου αστοχίας του κοντινότερου γείτονα προκύπτει η **Εξίσωση 24**.

$$\begin{aligned} & \text{Relative Probability of N.N. Failure Type (\%)} \\ = & \frac{W * \text{Rel. Failure Type Frequency of N.N. Failure Type} * \text{Relative Failure Probability (\%)}}{\Sigma(\text{Rel. Failure Type Frequency of All Possible Types})} \end{aligned}$$

Εξίσωση 24

Και αναλύοντας το άθροισμα των σχετικών συχνοτήτων οι Εξισώσεις **23** και **24**, μετασχηματίζονται αντιστοίχως στις Εξισώσεις **25** και **26** που ακολουθούν:

$$\begin{aligned} & \text{Relative Probability of Other Possible Failure Types (\%)} = \\ & \frac{\text{Rel. Failure Type Freq. of Other Possible Failure Type} * \text{Relative Failure Probability (\%)}}{W * \text{Rel. Fail. Type Freq. of N.N. Failure Type} + \Sigma(\text{Rel. Fail. Type Freq. of Other Possible Types})} \end{aligned}$$

Εξίσωση 25

$$\begin{aligned} & \text{Relative Probability of N.N. Failure Type (\%)} = \\ & \frac{W * \text{Rel. Failure Type Freq. of N.N. Failure Type} * \text{Relative Failure Probability (\%)}}{W * \text{Rel. Fail. Type Freq. of N.N. Failure Type} + \Sigma(\text{Rel. Fail. Type Freq. of Other Possible Types})} \end{aligned}$$

Εξίσωση 26

Κατόπιν των ανωτέρω, για την υπό εξέταση σήραγγα του παραδείγματος και ορίζοντας τυχαία μία τιμή για τον συντελεστή βαρύτητας $W = 4$, προκύπτει σύμφωνα με την Εξίσωση 25:

$$\text{Rel. Probability of FAC failure type (\%)} = \frac{0,57 * 39,99}{4 * 0,34 + (0,57 + 0,07 + 0,02)} = 11,2\%$$

$$\text{Rel. Probability of CRW failure type (\%)} = \frac{0,07 * 39,99}{4 * 0,34 + (0,57 + 0,07 + 0,02)} = 1,35\%$$

$$\text{Rel. Probability of FAC/OSF failure type (\%)} = \frac{0,07 * 39,99}{4 * 0,34 + (0,57 + 0,07 + 0,02)} = 0,45\%$$

Και σύμφωνα με την Εξίσωση 26:

$$\text{Rel. Probability of OSF failure type (\%)} = \frac{4 * 0,34 * 39,99\%}{4 * 0,34 + (0,57 + 0,07 + 0,02)} = 26,9\%$$

Προς επαλήθευση, αθροίζονται οι ανωτέρω σχετικές πιθανότητες εμφάνισης κάθε τύπου αστοχίας και προκύπτει η συνολική σχετική πιθανότητα αστοχίας της υπό εξέταση σήραγγας, δηλ. $11,2\% + 1,35\% + 0,45\% + 26,9\% = 39,9\%$. Τα ίδια αποτελέσματα φαίνονται και στην **Εικόνα 34**, από την εφαρμογή του μοντέλου.

The Relative Failure Probability for each type of failure is given to the table below:

Failure_Type	Relative_Failure_Probability_for_Failure_Type_%
0 FAC	11.232159
1 OSF	26.957181
2 CRW	1.347859
3 FAC/OSF	0.449286

Εικόνα 34. Αποτελέσματα εφαρμογής μοντέλου για τον υπολογισμό της σχετικής πιθανότητας εμφάνισης συγκεκριμένου τύπου αστοχίας για μία υπό εξέταση σήραγγα.

7.2 Υπολογισμός Ακρίβειας Πρόβλεψης Τύπου Αστοχίας

Για τον υπολογισμό της ακρίβειας πρόβλεψης τύπου αστοχίας των μοντέλων, το dataset TFD_21 διαχωρίστηκε χρησιμοποιώντας δύο αναλογίες διαχωρισμού, σε δύο (2) επιμέρους σύνολα δεδομένων για κάθε αναλογία. Το πρώτο σύνολο, περιλαμβάνει το 90% ή το 80% των εγγραφών του TFD_21 και είναι το σύνολο δεδομένων εκπαίδευσης των μοντέλων (training set, TFD_21_TR), ενώ το δεύτερο σύνολο περιλαμβάνει το υπόλοιπο 10% ή το 20% των εγγραφών αντιστοίχως και είναι το σύνολο δεδομένων ελέγχου (test set, TFD_21_TS). Η αναλογίες διαχωρισμού του TFD_21 ορίστηκαν σε εννέα προς ένα (9/1) και οκτώ προς δύο (8/2), καθότι για μεγαλύτερες αναλογίες θεωρήθηκε ότι το πλήθος των εγγραφών του training set θα ήταν εξαιρετικά μικρό για την εκπαίδευση των μοντέλων. Επομένως, προέκυψε το TFD_21_TR που περιλαμβάνει ενενήντα πέντε (95) εγγραφές ή ογδόντα επτά (87) εγγραφές και το TFD_21_TS που περιλαμβάνει δώδεκα (12) εγγραφές ή είκοσι (20) εγγραφές αντιστοίχως.

Η επιλογή των δεδομένων του TFD_21_TS δεν ήταν τυχαία. Κατόπιν εφαρμογής του μοντέλου TFM1_A στο TFD_21, προέκυψαν τα τέσσερα (4) clusters. Από κάθε cluster επιλέχθηκαν τρία (3) ή πέντε (5) περιστατικά με τέτοιο τρόπο, ώστε να καλύπτεται (κατά το δυνατόν) όλο το εύρος τιμών που εμφανίζονται σε κάθε cluster, για κάθε χαρακτηριστικό. Ο λόγος που επιλέχθηκαν τα cluster που προκύπτουν από εφαρμογή του TFM1_A, ήταν επειδή από την εφαρμογή του TFM2_A, προκύπτει “Low Frequency” cluster με πέντε (5) εγγραφές και με την αφαίρεση τριών (3) ή πέντε (5) εγγραφών για χρήση στο test set, προκύπτει “Low Frequency” cluster με διαφορετικό πρότυπο χαρακτηριστικών, μετά την εκπαίδευση του μοντέλου στα εναπομείναντα περιστατικά. Αντιθέτως, με εξαγωγή τριών (3) ή πέντε (5) εγγραφών ανά cluster του TFM1_A, μετά την εκπαίδευση των μοντέλων στο TFD_21_TR, προκύπτουν συστάδες με το ίδιο πρότυπο χαρακτηριστικών που προκύπτουν από την εκπαίδευση με το TFD_21.

Ο υπολογισμός της ακρίβειας (*accuracy*), πραγματοποιήθηκε εισάγοντας κάθε εγγραφή του TFD_21_TS ως υπό εξέταση σήραγγα στο κάθε μοντέλο. Για την κάθε εγγραφή προκύπτει ένας τύπος αστοχίας που παρουσιάζει την μεγαλύτερη σχετική πιθανότητα εμφάνισης. Αν αυτός ο τύπος αστοχίας συμφωνεί με τον τύπο αστοχίας που παρουσιάστηκε στην πραγματικότητα, τότε η πρόβλεψη θεωρείται επιτυχής. Η ακρίβεια του μοντέλου δίνεται από το λόγο των επιτυχών προβλέψεων προς το σύνολο των προβλέψεων που πραγματοποιήθηκαν.

Μία πολύ σημαντική παράμετρος που λαμβάνεται υπόψη κατά την πρόβλεψη του τύπου αστοχίας για μία υπό εξέταση σήραγγα, είναι ο συντελεστής βαρύτητας W , που αναφέρθηκε στο Κεφάλαιο 7.1, ο οποίος εκφράζει τη βαρύτητα επιρροής του “κοντινότερου γείτονα” στον υπολογισμό της σχετικής πιθανότητας εμφάνισης τύπου αστοχίας μέσω των Εξιιώσεων 25 και 26. Η τιμή του συντελεστή W , ορίστηκε σε $W = 3$, καθότι με αυτή την τιμή τα μοντέλα επετύγχαναν την μεγαλύτερη ακρίβεια. Συγκεκριμένα, έγινε υπολογισμός τις ακρίβειας των μοντέλων για ακέραιες τιμές του W από 1 έως 5 (1, 2, 3, 4, 5) και για $W \geq 3$, η ακρίβεια των μοντέλων παρέμενε σταθερή. Παρά το γεγονός ότι η τιμή $W = 1$ υποδηλώνει ισότιμη βαρύτητα του cluster και του “κοντινότερου γείτονα” στην πρόβλεψη, το οποίο όπως αναφέρθηκε στο Κεφάλαιο 7.1, θεωρείται ότι δεν δύναται να ισχύει, ο υπολογισμός πραγματοποιήθηκε και με αυτή την τιμή, για την επαλήθευση αυτής της θεώρησης. Επομένως η τιμή $W = 3$, ορίστηκε ως η τελική τιμή του συντελεστή βαρύτητας, η οποία είναι αυτή που χρησιμοποιείται στην τελική μορφή του τρίτου τμήματος των μοντέλων. Η τιμή αυτή υποδηλώνει πως στα μοντέλα που αναπτύχθηκαν, η επιρροή του “κοντινότερου γείτονα”, ως προς τον υπολογισμό των σχετικών πιθανοτήτων εμφάνισης συγκεκριμένο τύπου αστοχίας, θεωρείται τρεις (3) φορές μεγαλύτερη από αυτή του cluster στο οποίο εμπίπτει μια υπό εξέταση σήραγγα. Στον Πίνακα 12, φαίνεται η ακρίβεια που επιτεύχθηκε για κάθε μοντέλο, αναλόγως της τιμής του συντελεστή βαρύτητας W και της αναλογίας διαχωρισμού, και στην Εικόνα 35 φαίνονται οι προβλέψεις των μοντέλων TFM1_C και TFM2_C για τις εγγραφές του TFD_21_TS, μέσω εκπαίδευσης στο TFD_21_TR, για $W = 4$ και αναλογία διαχωρισμού 9/1.

Αναλογία Διαχωρισμού (training/test)	Μοντέλο	Ακρίβεια (Accuracy) Μοντέλων (%)				
		$W=1$	$W=2$	$W=3$	$W=4$	$W=5$
90/10	TFM1 (K-Means)	0,42	0,75	0,75	0,75	0,75
	TFM2 (K-Prototypes)	0,67	0,75	0,92	0,92	0,92
80/20	TFM1 (K-Means)	0,40	0,70	0,70	0,70	0,70
	TFM2 (K-Prototypes)	0,45	0,45	0,60	0,60	0,60

Πίνακας 12. Ακρίβεια (accuracy) των μοντέλων TFM1 και TFM2, ως προς την πρόβλεψη τύπου αστοχίας σήραγγας, για τις διάφορες τιμές του συντελεστή βαρύτητας W .

TFM1 (K - Means) W=4																
INPUTS								RESULTS								
Project	Tunnel Use	Cutting Technique	Length	Diameter	Overburden	Rock/ Soil	Actual Failure Type	Nearest Neighbor Project	Relative Failure Probability for Failure Type (%)					Predicted Failure Type		
									FAC	OSF	CRW	FAC/OSF	FIR		LS	OTH
Thirre Tunnel	Road	NATM	5500	12.6	650	Rock	OSF	Mingyazi tunnel	11.57	25.91	1.39	0.46	0.00	0.00	0.00	OSF
Jinping II Hydropower Station headr 3	Hydro	TBM	1980	12.4	2500	Rock	OSF	Jinping II Hydropower Station drainage	4.86	21.87	1.21	0.00	0.00	0.00	0.00	OSF
Jinping II Hydropower Station headr 4	Hydro	D&B	2300	13	2500	Rock	OSF	Jinping II Hydropower Station headr 2	7.88	17.64	0.94	0.31	0.00	0.00	0.00	OSF
Shimenya Tunnel	Road	NATM	7524	11	1300	Rock	FAC	Karawanken Road	30.53	4.27	0.92	0.30	0.00	0.00	0.00	FAC
Headrace tunnel at Gilgel Gibe II Hydro 1	Hydro	TBM	26000	8.1	650	Rock	FAC	Karahnjukar hydropower project, Karahnjukar	28.69	8.07	1.79	0.00	0.00	0.00	0.00	FAC
Ettendorfer Tunnel	Road	NATM/OPEN	750	10.8	2	Soil	OSF	M6 Motorway	9.36	32.08	0.00	0.00	0.00	1.34	0.00	OSF
Cologne North-South Metro Tram Line	Metro	C&C	3800	7.3	40	Soil	OSF	Habsburgtunnel, National Highway A3, Brugg	30.28	6.49	0.00	0.00	0.00	1.08	0.00	FAC
Metro Tunnel Line 3 Extension, Athens	Metro	TBM	5600	10	15	Rock	FAC	Metro Tunnel Line 3, Athens	31.19	8.77	1.95	0.00	0.00	0.00	0.00	FAC
Kowloon Southern Link Canton Road	Rail	TBM	1200	8	3	Soil	FAC	Kowloon Southern Link Salisbury Road	37.64	4.70	0.00	0.00	0.78	0.00	0.78	FAC
Gibei Railway	Rail	HAND	2210	9	150	Soil	FAC	Singapore MRT, Niccholson	7.87	26.97	0.00	0.00	0.00	1.12	0.00	OSF
Vereina Tunnel, Central Section, Klosters	Rail	TBM	19000	9.7	1550	Rock	OSF	pressure shaft, hydraulic power plant Cleuson-	6.19	27.84	1.55	0.00	0.00	0.00	0.00	OSF
Severomuisk railway tunnel, Baikal-Amur Line	Rail	TBM	15343	4.56	1000	Rock	FAC	pressure shaft, hydraulic power plant Cleuson-	6.74	30.35	1.69	0.00	0.00	0.00	0.00	OSF
								Accuracy					0.73			
TFM2 (K - Prototypes) W=4																
INPUTS								RESULTS								
Project	Tunnel Use	Cutting Technique	Length	Diameter	Overburden	Rock/ Soil	Actual Failure Type	Nearest Neighbor Project	Relative Failure Probability for Failure Type (%)					Predicted Failure Type		
									FAC	OSF	CRW	FAC/OSF	FIR		LS	OTH
Thirre Tunnel	Road	NATM	5500	12.6	650	Rock	OSF	Mingyazi tunnel	9.26	23.40	0.49	0.00	0.00	0.49	0.00	OSF
Jinping II Hydropower Station headr 3	Hydro	TBM	1980	12.4	2500	Rock	OSF	Jinping II Hydropower Station drainage	2.47	19.77	1.23	0.00	0.00	0.00	0.00	OSF
Jinping II Hydropower Station headr 4	Hydro	D&B	2300	13	2500	Rock	OSF	Jinping II Hydropower Station headr 2	4.10	21.85	2.73	0.00	0.00	0.00	0.00	OSF
Shimenya Tunnel	Road	NATM	7524	11	1300	Rock	FAC	Karawanken Road	28.42	4.62	0.35	0.00	0.00	0.35	0.00	FAC
Headrace tunnel at Gilgel Gibe II Hydro 1	Hydro	TBM	26000	8.1	650	Rock	FAC	Karahnjukar hydropower project, Karahnjukar	21.63	2.70	0.78	0.39	0.00	0.00	0.00	FAC
Ettendorfer Tunnel	Road	NATM/OPEN	750	10.8	2	Soil	OSF	M6 Motorway	5.24	14.34	0.28	0.00	0.00	0.28	0.00	OSF
Cologne North-South Metro Tram Line	Metro	C&C	3800	7.3	40	Soil	OSF	Singapore MRT, Niccholson	7.27	16.97	1.82	0.00	0.00	0.00	0.00	OSF
Metro Tunnel Line 3 Extension, Athens	Metro	TBM	5600	10	15	Rock	FAC	Metro Tunnel Line 3, Athens	23.74	3.84	0.35	0.00	0.35	0.00	0.35	FAC
Kowloon Southern Link Canton Road	Rail	TBM	1200	8	3	Soil	FAC	Kowloon Southern Link Salisbury Road	23.33	3.77	0.34	0.00	0.34	0.00	0.34	FAC
Gibei Railway	Rail	HAND	2210	9	150	Soil	FAC	Channel Tunnel Rail Link	10.12	5.95	0.59	0.00	0.59	0.00	0.59	FAC
Vereina Tunnel, Central Section, Klosters	Rail	TBM	19000	9.7	1550	Rock	OSF	Perthus tunnel	8.57	20.00	2.14	0.00	0.00	0.00	0.00	OSF
Severomuisk railway tunnel, Baikal-Amur Line	Rail	TBM	15343	4.56	1000	Rock	FAC	Perthus tunnel	7.27	15.66	1.12	0.56	0.00	0.00	0.00	OSF
								Accuracy					0.92			

Εικόνα 35. Αποτελέσματα πρόβλεψης τύπου αστοχίας σήραγγας για τα μοντέλα TFM1 και TFM2 ($W = 4$, αναλογία διαχωρισμού 90/10).

Όπως φαίνεται στον Πίνακα 12, η αύξηση της τιμής του W , επέφερε αύξηση στην ακρίβεια πρόβλεψης τύπου αστοχίας. Ο “κοντινότερος γείτονας”, όπως προαναφέρθηκε, αποτελεί το περιστατικό - σημείο το οποίο απέχει τη μικρότερη ευκλείδεια απόσταση από το σημείο που εκφράζει την υπό εξέταση σήραγγα. Το γεγονός ότι η ακρίβεια αυξάνεται με την αύξηση επιρροής του “κοντινότερου γείτονα”, υποδηλώνει πως η βαρύτητα επιρροής ενός οποιοδήποτε σημείου – περιστατικού στην πρόβλεψη, αυξάνεται όσο μειώνεται η απόσταση αυτού από το εξεταζόμενο σημείο. Η παρατήρηση αυτή παραπέμπει στην παρεμβολή σταθμισμένης αντίστροφης απόστασης (Inverse Distance Weighted Interpolation, IDW) [33], βασική υπόθεση της οποίας είναι ότι κάθε γνωστό σημείο έχει μια τοπική επιρροή που μικραίνει όσο απομακρυνόμαστε από αυτό [34]. Επομένως, ένα επόμενο στάδιο έρευνας και πιθανώς βελτίωσης του μοντέλου TFM1 (K-Means), θα μπορούσε να περιλαμβάνει την εξέταση της ακρίβειας πρόβλεψης τύπου αστοχίας, μέσω εφαρμογής της μεθόδου IDW, για τις διάφορες τιμές της παραμέτρου (εκθέτη) “ u ” που χρησιμοποιεί η μέθοδος.

8 Συμπεράσματα

Συνοψίζοντας, στην παρούσα εργασία χρησιμοποιήθηκαν δεδομένα που περιλαμβάνουν χαρακτηριστικά σηράγγων, κατά την κατασκευή των οποίων παρουσιάστηκαν σημαντικές αστοχίες, με σκοπό την ομαδοποίηση των περιστατικών αυτών προς εξεύρεση πιθανόν προτύπων (μοτίβων) χαρακτηριστικών που οδηγούν σε ορισμένο κίνδυνο αστοχίας, αλλά και την αξιοποίηση αυτών των αποτελεσμάτων για την εξέταση πιθανών σεναρίων κατασκευής σηράγγων. Στο παρόν κεφάλαιο ερμηνεύονται τα αποτελέσματα των μοντέλων ανά τμήμα (A – συσταδοποίηση, B – εκτίμηση κινδύνου αστοχίας, C – πρόβλεψη τύπου αστοχίας) και εξάγονται συμπεράσματα. Επιπλέον, αναφέρονται τα προβλήματα που προέκυψαν κατά την εκπόνηση της εργασίας και παρατίθενται προτάσεις για την βελτίωση της αποτελεσματικότητας των μοντέλων αλλά και την περεταίρω έρευνα του θέματος.

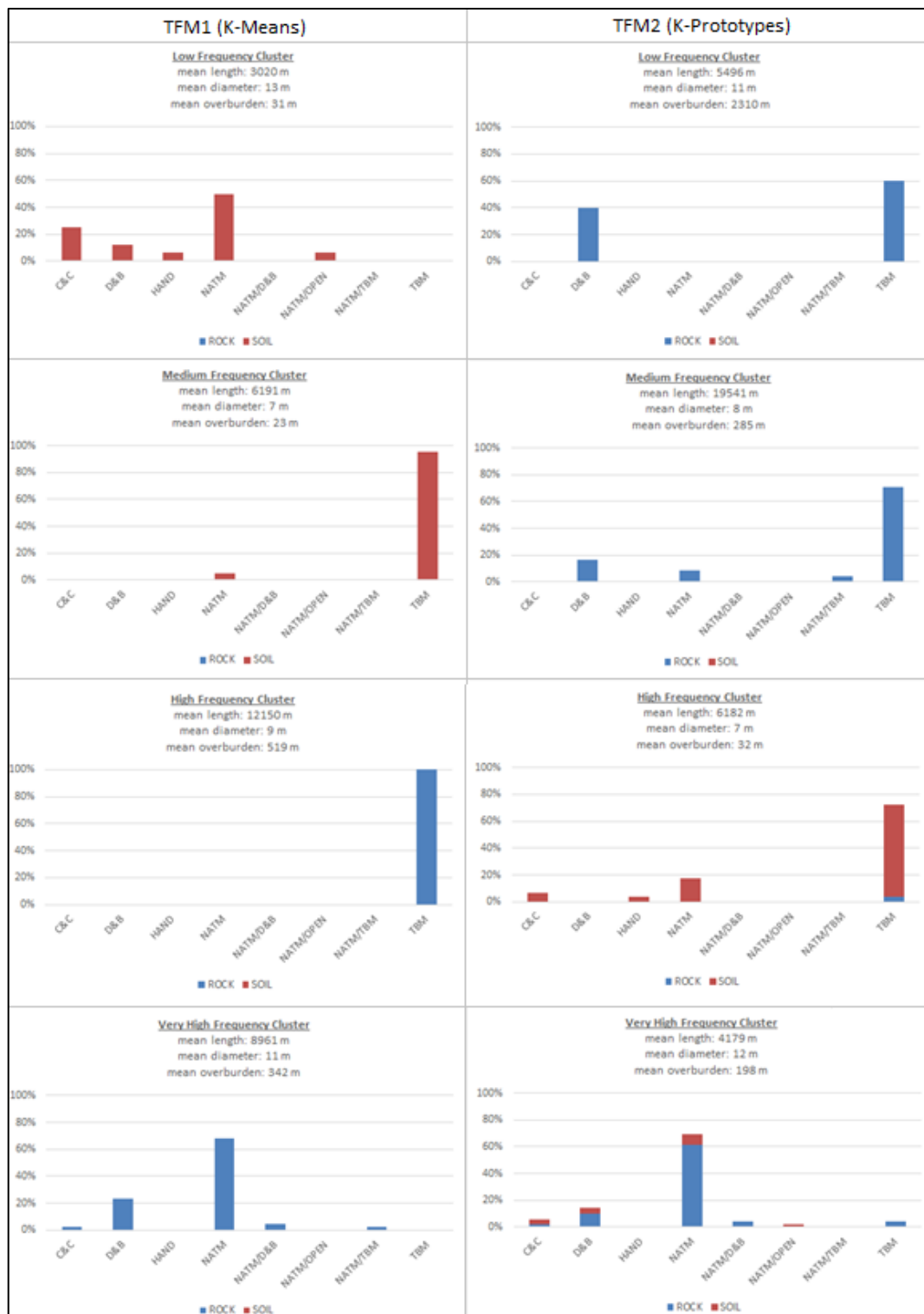
8.1 Συμπεράσματα TFM1_A, TFM2_A

Η διαμόρφωση των συστάδων (clusters), ήταν αποτέλεσμα εφαρμογής δύο αλγορίθμων μηχανικής μάθησης, του K-Means και του K-Prototypes, οι οποίοι ενσωματώθηκαν αντιστοίχως σε δύο μοντέλο με διαφορετική αρχιτεκτονική, το TFM1 και το TFM2. Η σήμανση (labeling) των clusters, υποδεικνύει τη σχετική συχνότητα (relative frequency) εμφάνισής τους στο dataset TFD_21, δηλαδή προκύπτει από τον λόγο των εγγραφών που περιλαμβάνονται στο cluster προς το σύνολο των εγγραφών. Στον Πίνακα 13 φαίνεται η σχετική συχνότητα εμφάνισης των cluster που προέκυψαν από την εφαρμογή του κάθε μοντέλου.

<i>Cluster Label</i>	<i>Cluster's Relative Frequency</i>	
	<i>TFM1 (K-Means)</i>	<i>TFM2 (K-Prototypes)</i>
Low Frequency	15,0 %	4,7 %
Medium Frequency	19,6 %	22,4 %
High Frequency	21,5 %	27,1 %
Very High Frequency	43,9 %	45,8 %

Πίνακας 13. Σχετική συχνότητα (relative frequency) εμφάνισης των cluster που προέκυψαν ανά μοντέλο και ποσοστιαία συμφωνία εγγραφών.

Στην **Εικόνα 36** απεικονίζονται διαγραμματικά οι τιμές των σημαντικών, σύμφωνα με την PCA (βλ. **Κεφ. 4.4.2**), χαρακτηριστικών ανά cluster και στον **Πίνακα 14** φαίνονται τα πιθανά πρότυπα (μοτίβα) χαρακτηριστικών που προκύπτουν από την ερμηνεία των αποτελεσμάτων.



Εικόνα 36. Διαγραμματική απεικόνιση χαρακτηριστικών cutting technique και rock soil ανά cluster (αριστερά TFM1, δεξιά TFM2).

<i>Feature</i>	<i>Low Frequency</i>	<i>Medium Frequency</i>	<i>High Frequency</i>	<i>Very High Frequency</i>
	<i>TFM1 (K-Means)</i>			
Cutting Technique	Conventional Methods	TBM	TBM	Conventional Methods
Length	< 7 km	< 7km	> 7km	> 7km
Diameter	> 10 m	< 10 m	< 10 m	> 10 m
Overburden	< 40 m	< 40 m	> 40 m	> 40 m
Rock/Soil	Soil	Soil	Rock	Rock
<i>TFM2 (K-Prototypes)</i>				
Cutting Technique	All methods	TBM	TBM	Conventional Methods
Length	< 2,5km	> 8km	< 8km	All lengths
Diameter	> 7m	< 10 m	< 10 m	> 10 m
Overburden	> 2000 m	> 40 m	< 40 m	All depths
Rock/Soil	Rock	Rock	Soil	Rock

Πίνακας 14. Πρότυπα (μοτίβα) χαρακτηριστικών που προκύπτουν από την διαμόρφωση των cluster.

Παρατηρώντας τα στοιχεία του Πίνακα 14 και τα διαγράμματα της [Εικόνας 36](#), συμπεραίνουμε ότι το πρότυπο του cluster “Very High Frequency” είναι κατ’ ουσία το ίδιο και στις δύο περιπτώσεις εφαρμογής, με τη διαφορά ότι μέσω του TFM1 το πρότυπο προκύπτει ως πιο “περιοριστικό”, ως προς τις τιμές των χαρακτηριστικών “length” και “overburden”. Επομένως, εφόσον από την ερμηνεία των αποτελεσμάτων και των δύο μοντέλων προκύπτει το ίδιο πρότυπο, αυτό αξιολογείται ως αξιόπιστο πρότυπο. Λαμβάνοντας υπόψη και την υψηλή σχετική συχνότητα εμφάνισης του cluster (43,9% και 45,8%), καταλήγουμε στο συμπέρασμα ότι οι σήραγγες που ακολουθούν αυτό το πρότυπο χαρακτηριστικών, εμπίπτουν στην κατηγορία υψηλού κινδύνου.

Επίσης, παρατηρούμε πως τα πρότυπα των cluster “High Frequency” και “Medium Frequency” συμφωνούν εναλλάξ για τους δύο αλγόριθμους. Για το λόγο αυτό, έγινε εξέταση συμφωνίας των εγγραφών που περιλαμβάνονται στα “Medium” και “High Frequency” clusters του TFM1 με αυτές που περιλαμβάνονται αντιστοίχως στα “High” και “Medium Frequency” clusters του TFM2. Ως αποτέλεσμα προέκυψε ότι οι εγγραφές του TFM1 “Medium Frequency” cluster περιλαμβάνονται σε ποσοστό 100% στο TFM2 “High Frequency” cluster και οι εγγραφές του TFM1 “High Frequency” cluster περιλαμβάνονται σε ποσοστό 73,9% στο TFM2 “Medium Frequency” cluster. Επομένως, λαμβάνοντας υπόψη

τα υψηλά ποσοστά συμφωνίας εγγραφών, αλλά και το γεγονός ότι τα συγκεκριμένα cluster παρουσιάζουν περίπου την ίδια συχνότητα εμφάνισης (από 19% έως 27% για κάθε μοντέλο), συμπεραίνουμε ότι τα δύο πρότυπα που προκύπτουν είναι αξιόπιστα και αφορούν στην ίδια κατηγορία κινδύνου, την μέση κατηγορία. Σε αυτή την κατηγορία εμπίπτουν σήραγγες που παρουσιάζουν οποιοδήποτε από τα δύο αυτά πρότυπα.

Όσον αφορά στο cluster “Low Frequency”, στην περίπτωση εφαρμογής του αλγορίθμου TFM1 προκύπτει ένα πρότυπο το οποίο δεν δύναται να εκφραστεί από τα πρότυπα των υπόλοιπων cluster του TFM1. Αντιθέτως, μέσω εφαρμογής του TFM2, το πρότυπο του “Low Frequency” cluster εκφράζεται συνδυαστικά από τα πρότυπα των “Medium” και “Very High Frequency” clusters. Γνωρίζουμε ότι στο TFM2 “Low Frequency” cluster περιλαμβάνονται πέντε εγγραφές, οι οποίες παρουσιάζουν ακραίες τιμές του χαρακτηριστικού “overburden”. Λόγω της εξαιρετικά υψηλής “απόστασης” που υπολογίζεται από τον αλγόριθμο K-Prototypes, διαμορφώνεται αυτό το cluster του οποίου όμως οι εγγραφές θα μπορούσαν να συμπεριληφθούν σε άλλα TFM2 clusters. Το ίδιο δεν συμβαίνει μέσω του αλγορίθμου K-Means (TFM1), διότι η απόσταση αυτή περιορίζεται σημαντικά λόγω του μετασχηματισμού των περιγραφικών μεταβλητών σε αριθμητικές. Εκ του αποτελέσματος συμπεραίνουμε ότι ενδεχομένως αν είχαν αφαιρεθεί αυτές οι εγγραφές από το dataset, θα είχαμε εμφάνιση των ίδιων cluster με τα ίδια πρότυπα μέσω εφαρμογής και των δύο μοντέλων. Επομένως, ως αξιόπιστο θεωρείται μόνο το πρότυπο του TFM1 “Low Frequency” cluster, το οποίο, λαμβάνοντας υπόψη και τη χαμηλή συχνότητα εμφάνισης του cluster, αφορά στην χαμηλή κατηγορία κινδύνου αστοχίας σήραγγας.

Κατόπιν των ανωτέρω, συνδυάζοντας τα αποτελέσματα συσταδοποίησης των δύο μοντέλων, καταλήγουμε στην ύπαρξη τεσσάρων (4) προτύπων χαρακτηριστικών τα οποία συσχετίζονται με τρεις (3) κατηγορίες κινδύνου όπως φαίνεται στον Πίνακα 15.

<i>Feature</i>	<i>Patterns</i>			
	<i>Low Risk Cat.</i>	<i>Medium Risk Category</i>		<i>High Risk Cat.</i>
Cutting Technique	Conventional Methods	TBM	TBM	Conventional Methods
Length	< 7 km	< 7km	> 8km	> 7km
Diameter	> 10 m	< 10 m	< 10 m	> 10 m
Overburden	< 40 m	< 40 m	> 40 m	> 40 m
Rock/Soil	Soil	Soil	Rock	Rock

Πίνακας 15. Πρότυπα (μοτίβα) χαρακτηριστικών που συσχετίζονται με ορισμένη κατηγορία κινδύνου (μέσω αξιολόγησης - συνδυασμού προτύπων TFM1 και TFM2).

Συνολικά, όσον αφορά στη λειτουργία συσταδοποίησης των μοντέλων, από το TFM1 προέκυψαν καλύτερα αποτελέσματα, από τα οποία διαφαίνονται σαφέστερα τα πρότυπα (patterns). Το πλεονέκτημα του TFM1 εντοπίζεται στην ικανότητα διαχείρισης στοιχείων με ακραίες τιμές αριθμητικών χαρακτηριστικών. Η ικανότητα αυτή είναι αποτέλεσμα της εργασίας μετασχηματισμού των περιγραφικών τιμών σε αριθμητικές, η οποία είναι απαραίτητο να εκτελείτε στο TFM1, προκειμένου να εφαρμοστεί ο αλγόριθμος που ενσωματώνει (K-Means), ο οποίος πραγματοποιεί υπολογισμό της τετραγωνικής ευκλείδειας απόστασης (βλ. Κεφ. 3.3.3.). Συγκεκριμένα, ο ορισμός των τιμών μετασχηματισμού για το χαρακτηριστικό “cutting technique”, έγινε με λογική κλιμάκωσης των τιμών από τις συμβατικές μεθόδους διάνοιξης προς τη μηχανοποιημένη όρυξη. Αυτό είχε ως αποτέλεσμα σήραγγες που διανοίχθηκαν με διαφορετική τεχνική, αλλά ως προς την κλιμάκωση τιμών γειτνιάζουσες (π.χ. τιμές: “D&B”, “NATM” = conventional), να παρουσιάζουν τιμή τετραγωνικής ευκλείδειας απόστασης μικρότερη του 1 (K-Means), έναντι του μέτρου ανομοιότητας (K-Prototypes) που λαμβάνει την τιμή 1 αν είναι διαφορετικές οι περιγραφικές τιμές απόστασης (βλ. Κεφ. 3.3.4.1). Η προαναφερθείσα διαπίστωση ερμηνεύει και τη σύσταση του TFM2 “Low Frequency” cluster. Η “απόσταση” των εγγραφών με εξαιρετικά υψηλή τιμή “overburden”, δεν μειώθηκε παρά την γειτνίαση των μεθόδων διάνοιξης με αυτές των λοιπών TFM2 clusters, και ως εκ τούτου διαμορφώθηκε μία ξεχωριστή και εξαιρετικά “απομακρυσμένη” συστάδα αποτελούμενη μόνο από αυτές τις εγγραφές.

8.2 Συμπεράσματα TFM1_B, TFM2_B

Για το τμήμα B των μοντέλων, αναπτύχθηκε μία μεθοδολογία εκτίμησης κινδύνου υπό εξέταση σήραγγας, η οποία στηρίζεται στις τεχνικές που χρησιμοποιεί ο αλγόριθμος του κάθε μοντέλου και η οποία προσομοιάζει τη λειτουργία του αλγορίθμου K-Nearest Neighbors. Η μεθοδολογία κωδικοποιήθηκε και ενσωματώθηκε στα μοντέλα, έτσι ώστε να παρέχεται η δυνατότητα στον χρήστη να εισάγει τα χαρακτηριστικά μιας υπό εξέταση σήραγγας και να εξάγει αποτελέσματα που αφορούν στην δυνατότητα ένταξής της σε μία από τις κατηγορίες κινδύνου αστοχίας, στην ποσοστιαία ομοιότητα της με το cluster στο οποίο εμπίπτει, στην ποσοστιαία ομοιότητα της με τον “κοντινότερο γείτονα” και την σχετική πιθανότητα αστοχίας της.

Δεδομένου ότι για τους υπολογισμούς που πραγματοποιούνται από αυτό το τμήμα των μοντέλων, χρησιμοποιούνται τα αποτελέσματα συσταδοποίησης, αλλά και επειδή ως μέτρο γειτνίασης για την εύρεση του “κοντινότερου γείτονα” (nearest neighbor) μίας υπό εξέτασης σήραγγας χρησιμοποιείτε το αντίστοιχο κάθε φορά μέτρο που χρησιμοποιεί ο αλγόριθμος, η αποτελεσματικότητα αυτού του τμήματος των μοντέλων είναι όμοια με την αντίστοιχη του τμήματος A. Δηλαδή το μοντέλο TFM2, δεν θεωρείται αποτελεσματικό για την εξέταση σηράγγων που παρουσιάζουν ακραίες τιμές αριθμητικών χαρακτηριστικών. Επομένως, για συνήθεις περιπτώσεις σεναρίων κατασκευής σηράγγων, θεωρείται πως τα τμήματα B των μοντέλων είναι εξίσου αποτελεσματικά

Συγκεκριμένα, η εύρεση του “κοντινότερου γείτονα”, αποτελεί εξαιρετικά σημαντική λειτουργία καθότι δύναται να χρησιμοποιηθεί για την άντληση πληροφοριών που αφορούν σε οικονομικά και χρονικά στοιχεία του έργου και ακριβή υπολογισμό του κινδύνου υπό αυτούς τους όρους. Όσον αφορά στον υπολογισμό της σχετικής πιθανότητας αστοχίας, επισημαίνεται πως το εργαλείο που αναπτύχθηκε για την παρούσα εργασία, αποτελεί εργαλείο σύγκρισης σεναρίων κατασκευής σηράγγων για χρήση σε πρώιμο στάδιο της έρευνας για την υλοποίηση ενός έργου. Για αυτή τη χρήση, ο υπολογισμός σχετικής πιθανότητας αστοχίας αποδίδει αποτελεσματικά τη διαφορά που εντοπίζεται ανάμεσα σε σενάρια κατασκευής σηράγγων που εμπίπτουν στην ίδια κατηγορία κινδύνου αστοχίας. Δηλαδή για σήραγγες που τηρούν το ίδιο πρότυπο χαρακτηριστικών, αυτή η λειτουργία μπορεί να οδηγήσει στην επιλογή του ασφαλέστερου σεναρίου, καθότι συνυπολογίζει τα αποτελέσματα ομοιότητας με τα στοιχεία εντός του cluster (similarity to cluster’s centroid, similarity to nearest neighbor). Επιπλέον, λαμβάνοντας υπόψη το γεγονός ότι μία υπό εξέταση σήραγγα υπάρχει περίπτωση να μην τηρεί απολύτως κάποιο από τα πρότυπα, δηλαδή να μην παρουσιάζει συμφωνία σε πέντε από τα πέντε χαρακτηριστικά, ενδείκνυται η χρήση και των δύο μοντέλων σε κάθε περίπτωση και διασταύρωση των αποτελεσμάτων.

8.3 Συμπεράσματα TFM1_C, TFM2_C

Για αυτό το σκέλος της εργασίας αναπτύχθηκε το τρίτο μέρος των μοντέλων, προκειμένου να εξεταστεί η δυνατότητα πρόβλεψης συγκεκριμένου τύπου αστοχίας που μπορεί να εμφανίσει μία υπό εξέταση σήραγγα, ελέγχοντας την ακρίβεια πρόβλεψης των μοντέλων μέσω εξέτασης περιστατικών αστοχίας με γνωστό αποτέλεσμα. Η πρόβλεψη βασίζεται στην συχνότητα εμφάνισης ενός τύπου αστοχίας εντός του cluster στο οποίο

εμπίπτει η υπό εξέταση σήραγγα και στον τύπο αστοχίας που εμφάνισε ο “κοντινότερος γείτονας”. Σκοπός υλοποίησης αυτού του σκέλους της εργασίας, δεν ήταν μόνο η αξιολόγηση της δυνατότητας πρόβλεψης, αλλά και ο προσδιορισμός της βαρύτητας επιρροής του κοντινότερου γείτονα στην πρόβλεψη.

Κρίνοντας από την ακρίβεια που επιτεύχθηκε από τα μοντέλα, 75% για το TFM1 και 92% για το TFM2 για αναλογία διαχωρισμού 9/1, και 70% για το TFM1 και 60% για το TFM2 για αναλογία διαχωρισμού 8/2, η ικανότητα πρόβλεψης θα μπορούσε να θεωρηθεί ικανοποιητική και για τα δύο μοντέλα. Λαμβάνοντας όμως υπόψη το γεγονός ότι για την πρόβλεψη δεν λαμβάνονται υπόψη γεωτεχνικές παράμετροι, καθότι δεν υπάρχουν στοιχεία για αυτές, η αξιοπιστία της πρόβλεψης δεν μπορεί να θεωρηθεί υψηλή.

Πάραυτα, σημαντική θεωρείται η παρατήρηση αύξησης στην ακρίβεια πρόβλεψης των δύο μοντέλων και για τις δύο αναλογίες διαχωρισμού, μέσω αύξησης του συντελεστή βαρύτητας W (βλ. Κεφ. 7.2., Πίνακας 12), ο οποίος εκφράζει τη βαρύτητα επιρροής του “κοντινότερου γείτονα” στον υπολογισμό της σχετικής πιθανότητας εμφάνισης συγκεκριμένου τύπου αστοχίας. Η απόδοση συνολικά και των δύο μοντέλων, επιτεύχθηκε για συντελεστή βαρύτητας $W = 3$, το οποίο υποδεικνύει αρκετά πιο έντονη επιρροή του “κοντινότερου γείτονα” στην πρόβλεψη, σε σχέση με την επιρροή του cluster στο οποίο εμπίπτει η προς εξέταση σήραγγα. Επισημαίνεται όμως, ότι αυτό αποτελεί απλώς μία ένδειξη επιρροής και ότι ο υπολογισμός του συντελεστή θα ήταν δυνατός μόνο αν πρόκυπταν αξιόπιστα αποτελέσματα πρόβλεψης υψηλής ακρίβειας.

8.4 Προβλήματα – Προτάσεις Περαιτέρω Έρευνας

Τα προβλήματα που παρουσιάστηκαν κατά την εκπόνηση της εργασίας αφορούν στη βάση δεδομένων που χρησιμοποιήθηκε για τη διαμόρφωση του dataset TFD_21 και στην δυνατότητα προεπεξεργασίας της, εντός του διαθέσιμου χρόνου εκπόνησης της εργασίας. Βασικότερο πρόβλημα, το οποίο δεν αντιμετωπίστηκε επαρκώς, ήταν η εξαιρετικά χαμηλή πλήρωση τιμών για ορισμένα χαρακτηριστικά των σηράγγων. Η πλήρωση χαμένων τιμών που επιτεύχθηκε ήταν χαμηλή (κατά μέσο όρο 4%) και ως εκ τούτου δεν ήταν δυνατή η διαμόρφωση dataset, με περισσότερα χαρακτηριστικά ή/και εγγραφές, από την ανάλυση του οποίου θα εξάγονταν πιο ακριβή αποτελέσματα. Η ύπαρξη εσφαλμένων τιμών ήταν ένα πρόβλημα που αντιμετωπίστηκε επαρκώς, όχι όμως πιθανότατα στο μέγιστο δυνατό βαθμό.

Όσον αφορά στη δημιουργία των μοντέλων δεν αντιμετωπίστηκαν προβλήματα, εντοπίστηκαν όμως σημεία τα οποία επιδέχονται βελτίωσης και ενδεχομένως θα αύξαναν την αποτελεσματικότητα των μοντέλων. Μια πιθανή συνέχεια της εργασίας, θα μπορούσε να περιλαμβάνει μετασχηματισμό των τιμών του χαρακτηριστικού “cutting technique” για το μοντέλο TFM2 και εκ νέου σύγκριση των συστάδων και των προτύπων που προκύπτουν. Επιπλέον, ενδιαφέρον ενδέχεται να παρουσιάζαν τα αποτελέσματα συσταδοποίησης στην περίπτωση σήμανσης των συστάδων, όχι βάσει συχνότητας εμφάνισης του cluster (number of incidents in cluster / number of incidents in dataset), αλλά βάσει εμφάνισης αστοχιών ανά χιλιόμετρο κατασκευής $[(\text{number of incidents in cluster} / \text{sum of lengths in cluster}) / (\text{number of incidents in dataset} / \text{sum of lengths in dataset})]$. Όσον αφορά στην πρόβλεψη τύπου αστοχίας, ένα επόμενο στάδιο έρευνας και πιθανώς βελτίωσης του μοντέλου TFM1 (K-Means), θα ήταν η εξέταση της ακρίβειας πρόβλεψης τύπου αστοχίας, μέσω κωδικοποίησης και εφαρμογής της μεθόδου παρεμβολής IDW (Inverse Distance Weighted Interpolation) [33], για τις διάφορες τιμές της παραμέτρου (εκθέτη) “u” που χρησιμοποιεί η μέθοδος και σε περίπτωση επίτευξης μεγαλύτερης ακρίβειας, ενσωμάτωσης του κώδικα στο μοντέλο.

Λαμβάνοντας υπόψη όλα τα ανωτέρω, θεωρείται ότι η περαιτέρω έρευνα του αντικειμένου θα έπρεπε να στραφεί κατά κύριο λόγο προς την βελτίωση της βάσης δεδομένων και δευτερευόντως προς την βελτίωση των μοντέλων.

9 Βιβλιογραφία

- [1] Α. Σοφιανός, Μέτρα Στήριξης Σηράγγων, Αθήνα: Εθνικό Μετσόβιο Πολυτεχνείο, 2018.
- [2] K. Turner, «Challenges and Trends for Geological Modelling and Visualisation,» *Bulletin of Engineering Geology and the Environment*, pp. 109-127, 65(2), 2005.
- [3] V. Romero και J. Stolz, «Cost Estimating for Underground Transit: Too Dangerous to "Guestimate",» *The Capital Projects Process*, 2015.
- [4] B. Flyvbjerg, M. Skamris και S. Buhl, «What Causes Cost Overrun in Transport Infrastructure,» *Transport Reviews*, pp. 3-18, 24(1), 2004.
- [5] T. Konstantis, S. Konstantis και P. Spyridis, «Tunnel Losses: Causes, Impacts, Trends and Risk Management,» σε *ITA World Tunnel Congress*, San Fransisco, 2016.
- [6] C. Aldred, «Series of Tunnel Losses Prompts Risk Management Code,» *Business Insurance*, 2007.
- [7] P. Perello, «Estimate of the Reliability in Geological Forecasts for Tunnels: Towards a Structured Approach,» *Rock Mech. Rock Eng.*, pp. 671-694, 44(6), 2011.
- [8] C. Paraskevopoulou και G. Boutsis, «Cost Overruns in Tunnelling Projects: Investigating the Impact of Geological and Geotechnical Uncertainty Using Case Studies,» *Infrastructures*, 5(9), 73, 2020.
- [9] R. Butnitz, G. Apostolakis, D. Boore, L. Cluff, K. Coppersmith, C. Cornell και P. Morris, «Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts,» U.S. Nuclear Regulatory Commission , Washington, 1997.
- [10] G. Venturini, N. Bianchi και M. Diederichs, «How to Quantify the Reliability of a Geological and Geotechnical Reference Model in Underground Projects,» *RETC*, pp. 525-537, 2019.
- [11] T. G. Carter, «Prediction and Uncertainties in Geological Engineering and Rock Mass Characterization Assessments,» σε *Quarto Ciclo di Conferenze di Meccanica e Ingegneria delle Rocce*, Torino, 1992.
- [12] Ι. Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας και Η. Σακελλαρίου, Τεχνητή Νοημοσύνη, 3η Έκδοση, Θεσσαλονίκη: Γκιούρδας Β. Εκδοτική, 2006.
- [13] Μ. Βαζιργιάννης και Μ. Χαλκίδη, Εξόρυξη Γνώσης από Βάσεις Δεδομένων, 2η Έκδοση, Αθήνα: Τυπωθήτω - Γιώργος Δαρδάνος, 2005.

- [14] A. Jain και R. Dubes, Algorithms for Clustering Data, New Jersey: Prentice Hall, 1988.
- [15] L. Kayfman και P. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, New Jersey: John Wilen and Sons, 1990.
- [16] S. Theodoridis και K. Koutroumpas, Pattern Recognition, London: Academic Press, 1999.
- [17] R. Rezaee, B. Lelieveldt και J. H. Reiber, «A New Cluster Validity Index for the Fuzzy C-Mean,» *Pattern Recognition Letters*, pp. 237-246, March 1998.
- [18] J. MacQueen, «Some Methods for Classification and Analysis of Multivariate Observations,» σε *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967.
- [19] Z. Huang, «Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values,» *Data Mining and Knowledge Discovery*, pp. 283-304, September 2, 1998.
- [20] J. Hedge και B. Rokseth, «Applications of machine learning methods for engineering risk assessment – A review,» *Safety Science*, 122, 2020.
- [21] A. Benardos και D. Kaliampakos, «Modelling TBM Performance with Artificial Neural Networks,» *Tunnelling and Underground Space Technology*, pp. 597-605, 19, 2004.
- [22] A. Benardos, «Artificial Intelligence in Underground Development: A Study of TBM Performance,» *Underground Spaces*, pp. 21-32, 2008.
- [23] R. Sousa και H. Einstein, «Risk Analysis During Tunnel Construction Using Bayesian Networks: Porto Metro Case Study,» *Tunnelling and Underground Space Technology*, pp. 86-100, 27(1), 2012.
- [24] L. Y. Ding και C. Zhou, «Development of Web-Based System for Safety Risk Early Warning in Urban Metro Construction,» *Automation in Construction*, pp. 45-55, 34, 2013.
- [25] K.-C. Hyun, S. Min, H. Choi, J. Park και I.-M. Lee, «Risk Analysis Using Fault-Tree Analysis (FTA) and Analytic Hierarchy Process (AHP) Applicable to Shield TBM Tunnels,» *Tunnelling and Underground Space Technology*, pp. 121-129, 49, 2015.
- [26] R. Mikaeli, S. S. Haghshenas και Z. Sedaghati, «Geotechnical Risk Evaluation of Tunneling Projects Using Optimization Techniques (Case Study: The Second Part of Emamzade Hashem Tunnel),» *Natural Hazards*, pp. 1099-1113, 97, 2019.
- [27] M. C. Dallavalle, «Assessing The Failure Potential of Underground Works and the Impact of Delays and Costs on Insurability,» *MSc Engineering Geology, University of Leeds*, 2020.

- [28] Ε. Κύρκος, Επιχειρηματική Ευφυΐα και Εξόρυξη Δεδομένων: Ανακάλυψη Γνώσης για τη Λήψη Επιχειρηματικών Αποφάσεων, Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Κάλλιπος, 2015.
- [29] Π. Δημήτριος, Ανάλυση Πολυμεταβλητών Τεχνικών, Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Κάλλιπος, 2015.
- [30] F. S. Shie, M.-Y. Chen και Y.-S. Liu, «Prediction of Corporate Financial Distress: An Application of the America Banking Industry,» *Neural Computing and Applications*, pp. 1687-1696, 2012.
- [31] M.-Y. Chen, «Bankruptcy Prediction in Firms with Statistical and Intelligent Techniques and a Comparison of Evolutionary Computation Approaches,» *Computers & Mathematics with Applications*, pp. 4514-4524, 2011.
- [32] S. Alija, F. J. Torrijo και M. Quinta-Ferreira, «Study of The Unexpected Collapse of the Ampurdan Tunnel (Spain),» *Bulletin of Engineering Geology and the Environment*, pp. 451-463, 73(2), 2013.
- [33] D. Shepard, «A Two-Dimensional Interpolation Function for Irregularly-Spaced Data,» σε *23rd ACM National Conference*, 1968.
- [34] Π. Ηλιοπούλου, Γεωγραφική Ανάλυση, Αθήνα: Σύνδεσμος Ακαδημαϊκών Ελληνικών Βιβλιοθηκών, 2015.