



**ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ**

**ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ**

**«ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ- ΜΒΑ»**

---

**ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΗΝ ΑΞΙΟΛΟΓΗΣΗ  
ΑΜΟΙΒΑΙΩΝ ΚΕΦΑΛΑΙΩΝ**

Υπό

**ΒΑΣΙΛΕΙΟΣ ΚΟΥΤΣΟΡΙΝΑΚΗΣ**

ΑΜ: 2021019035

Επιβλέπων Καθηγητής: ΜΙΧΑΛΗΣ ΔΟΥΜΠΟΣ

**ΧΑΝΙΑ 2023**

## ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της μεταπτυχιακής διπλωματικής μου εργασίας, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους συνέβαλλαν στην εκπόνησή της και στην επιτυχή ολοκλήρωση των σπουδών μου.

Ευχαριστώ θερμά τον επιβλέπων καθηγητή μου κύριο Δούμπο Μιχάλη, για την εμπιστοσύνη που μου έδειξε εξ' αρχής, αναθέτοντάς μου το συγκεκριμένο θέμα, την επιστημονική του καθοδήγηση, τις υποδείξεις του, την επιμονή του, τη συμπαράστασή του, τη συνεχή του υποστήριξη, το αμείωτο ενδιαφέρον που έδειξε από την αρχή μέχρι το τέλος, καθώς και για τις γνώσεις που μου προσέφερε κατά την διάρκεια της μεταπτυχιακής μου πορείας.

Τέλος, θα ήθελα εκφράσω την ευγνωμοσύνη μου στην σύζυγό μου Τσαφαντάκη Μαρκέλλα και στα παιδιά μας Ειρήνη και Νίκο για όλη τη στήριξη, τη συμπαράσταση και την κατανόησή τους, καθ' όλη τη διάρκεια των σπουδών μου.

## ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ- ΣΚΟΠΟΣ ΕΡΓΑΣΙΑΣ.....	4
2. ΤΟ ΠΛΑΙΣΙΟ ΛΕΙΤΟΥΡΓΙΑΣ ΤΩΝ ΑΜΟΙΒΑΙΩΝ ΚΕΦΑΛΑΙΩΝ.....	6
2.1 Σύντομη αναδρομή δημιουργίας και ανάπτυξης των αμοιβαίων κεφαλαίων .....	6
2.2 Τι είναι αμοιβαίο κεφάλαιο .....	9
2.3 Κατηγορίες αμοιβαίων κεφαλαίων.....	11
2.4 Πλεονεκτήματα-μειονεκτήματα αμοιβαίων κεφαλαίων .....	16
3. ΜΕΘΟΔΟΙ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΗΝ ΑΞΙΟΛΟΓΗΣΗ ΑΜΟΙΒΑΙΩΝ ΚΕΦΑΛΑΙΩΝ .....	19
3.1 Το πλαίσιο της τεχνητής νοημοσύνης.....	19
3.2 Αναγκαιότητα ανάπτυξης μεθόδων Μηχανικής Μάθησης .....	22
3.2.1 Μη Επιτηρούμενη Μάθηση (Unsupervised Learning) .....	23
3.2.2 Ενισχυτική Μάθηση (Reinforcement Learning).....	24
3.2.3 Επιτηρούμενη Μάθηση (Supervised Learning).....	25
4. ΠΕΙΡΑΜΑΤΙΚΟ- ΕΡΕΥΝΗΤΙΚΟ ΣΚΕΛΟΣ .....	33
4.1 Εκκαθάριση- Προεπεξεργασία Δεδομένων .....	33
4.2 Αποτελέσματα και σχολιασμός.....	40
4.2.1 Σημαντικότητα μεταβλητών .....	40
4.2.2 Αποτελέσματα ταξινόμησης.....	43
5. ΣΥΜΠΕΡΑΣΜΑΤΑ .....	48
ΒΙΒΛΙΟΓΡΑΦΙΑ .....	50

## 1. ΕΙΣΑΓΩΓΗ- ΣΚΟΠΟΣ ΕΡΓΑΣΙΑΣ

Τα τελευταία 40 έτη η αλλαγή της επενδυτικής φιλοσοφίας που ακολουθείται τόσο παγκοσμίως όσο και από τους Έλληνες επενδυτές, έχει ευνοήσει την ανάπτυξη της αγοράς των αμοιβαίων κεφαλαίων. Η μεγαλύτερη ανάπτυξη παρουσιάζεται κυρίως στις Ηνωμένες Πολιτείες Αμερικής (ΗΠΑ) γεγονός που οφείλεται τόσο στη δομή της οικονομίας όσο και στην κουλτούρα των επενδυτών. Στην Ελλάδα η αγορά των αμοιβαίων κεφαλαίων παρουσιάζει αλματώδη αύξηση τα τελευταία 15 περίπου χρόνια, ως αποτέλεσμα ενός συνδυασμού νέων συνθηκών με τις οποίες ήρθε αντιμέτωπος ο Έλληνας επενδυτής κυρίως, λόγω της ύπαρξης σταθερού, ισχυρού νομίσματος με χαμηλά επιτόκια δανεισμού. Η επένδυση σε αμοιβαία κεφάλαια αποτελεί πλέον μία από τις πιο διαδεδομένες μορφές επενδύσεων που παρέχει πολλά πλεονεκτήματα συγκριτικά με τις παραδοσιακές μορφές επενδύσεων (όπως μετοχές, ομόλογα, προθεσμιακές καταθέσεις, κ.ά.) που θα αναλυθούν παρακάτω.

Η εργασία χωρίζεται σε 4 κεφάλαια. Το 1<sup>ο</sup> είναι μία εισαγωγή για τα αμοιβαία κεφάλαια και μία σύντομη περιγραφή της εργασίας. Στο 2<sup>ο</sup> κεφάλαιο αρχικά γίνεται μία σύντομη περιγραφή του τί είναι αμοιβαία κεφάλαια, μία σύντομη αναδρομή στην ιστορία των αμοιβαίων κεφαλαίων, στο πως ξεκίνησαν, πως αναπτύχθηκαν και ποιο είναι το μέγεθος της αγοράς των αμοιβαίων κεφαλαίων την παρούσα χρονική στιγμή τόσο σε τοπικό όσο και σε διεθνές επίπεδο. Στην συνέχεια γίνεται μία σύντομη περιγραφή στις κατηγορίες αμοιβαίων κεφαλαίων καθώς και στα πλεονεκτήματα και τα μειονεκτήματα της επένδυσης σε εταιρείες αμοιβαίων κεφαλαίων. Στο 3<sup>ο</sup> κεφάλαιο της εργασίας περιγράφονται συνοπτικά κάποιες από τις σημαντικότερες και δημοφιλέστερες μεθόδους τεχνητής νοημοσύνης και συστημάτων μηχανικής μάθησης που έχουν χρησιμοποιηθεί στην αξιολόγηση των αμοιβαίων κεφαλαίων, καθώς και σε μεθόδους μη επιτηρούμενης, ενισχυτικής και επιτηρούμενης μάθησης. Ιδιαίτερη έμφαση δόθηκε στην κατηγορία των μεθόδων επιτηρούμενης μάθησης, καθώς τέσσερις από αυτές ( μονότονη λογιστική παλινδρόμηση, XGBoost, Random Forest και μονότονη Random Forest) χρησιμοποιήθηκαν για το πειραματικό σκέλος της εργασίας. Η υλοποίηση των παραπάνω μεθόδων έγινε σε περιβάλλον R μέσω του πακέτου Caret. Στο 4<sup>ο</sup> κεφάλαιο καταγράφεται το πειραματικό σκέλος της εργασίας. Χρησιμοποιώντας ένα δείγμα αμοιβαίων κεφαλαίων προσπαθήσαμε μέσω εφαρμογής μεθόδων τεχνητής νοημοσύνης να προβλέψουμε την αξιολόγησή τους σε σχέση με την αξιολόγηση της εταιρείας Morningstar. Αρχικά αναφέρθηκε η πηγή των δεδομένων, καθώς και η διαδικασία προεπεξεργασίας τους. Έπειτα αναφέρθηκαν οι διαδικασίες ανάλυσης και μείωσης του αριθμού των μεταβλητών μελετώντας την συσχέτιση μεταξύ τους και την σημαντικότητά τους, πραγματοποιώντας δύο στατιστικούς

ελέγχους την ANOVA (Analysis of Variance) και του ελέγχου Kruskal-Wallis, για την κάθε κατηγορία ταξινόμησης. Οι κατηγορίες ταξινόμησης είναι η ταξινόμηση η αξιολόγηση των κεφαλαίων με βάση τον κίνδυνο, η αξιολόγηση των κεφαλαίων με βάση την απόδοσή τους και η συνολική αξιολόγηση των αμοιβαίων κεφαλαίων. Τα αποτελέσματα των παραπάνω στατιστικών ελέγχων οδήγησαν τελικά στην επιλογή 22 μεταβλητών που φαίνεται να μην έχουν υψηλή μεταξύ τους συσχέτιση και να είναι στατιστικά σημαντικές. Δηλαδή προσπαθήσαμε να καταλάβουμε με ποια μέθοδο η αξιολόγησή μας τείνει να προσεγγίσει την αξιολόγηση της Morningstar. Τέλος στο 5<sup>ο</sup> κεφάλαιο, που είναι και ο επίλογος, καταγράφονται κάποια συνοπτικά συμπεράσματα.

Σκοπός της εργασίας είναι να διαπιστώσουμε με ποια μέθοδο τεχνητής νοημοσύνης η αξιολόγηση των αμοιβαίων κεφαλαίων προσεγγίζει την αξιολόγηση της MorningStar, την οποία πολλοί επενδυτές λαμβάνουν υπόψιν τους για τις επενδυτικές τους κινήσεις.

## 2. ΤΟ ΠΛΑΙΣΙΟ ΛΕΙΤΟΥΡΓΙΑΣ ΤΩΝ ΑΜΟΙΒΑΙΩΝ ΚΕΦΑΛΑΙΩΝ

### 2.1 Σύντομη αναδρομή δημιουργίας και ανάπτυξης των αμοιβαίων κεφαλαίων

Τα τελευταία 40 έτη η αλλαγή της επενδυτικής φιλοσοφίας που ακολουθείται τόσο παγκοσμίως όσο και από τους Έλληνες επενδυτές, έχει ευνοήσει την ανάπτυξη της αγοράς των αμοιβαίων κεφαλαίων. Η μεγαλύτερη ανάπτυξη παρουσιάζεται κυρίως στις Ηνωμένες Πολιτείες Αμερικής (ΗΠΑ) γεγονός που οφείλεται τόσο στη δομή της οικονομίας όσο και στην κουλτούρα των επενδυτών. Στην Ελλάδα η αγορά των αμοιβαίων κεφαλαίων παρουσιάζει αλματώδη αύξηση τα τελευταία 15 περίπου χρόνια, ως αποτέλεσμα ενός συνδυασμού νέων συνθηκών με τις οποίες ήρθε αντιμέτωπος ο Έλληνας επενδυτής κυρίως, λόγω της ύπαρξης σταθερού, ισχυρού νομίσματος με χαμηλά επιτόκια δανεισμού. Η επένδυση σε αμοιβαία κεφάλαια αποτελεί πλέον μία από τις πιο διαδεδομένες μορφές επενδύσεων που παρέχει πολλά πλεονεκτήματα συγκριτικά με τις παραδοσιακές μορφές επενδύσεων (όπως μετοχές, ομόλογα, προθεσμιακές καταθέσεις, κ.ά.) που θα αναλυθούν παρακάτω.

Αν και πολλοί εικάζουν ότι τα αμοιβαία κεφάλαια ως μορφή επένδυσης ξεκίνησαν από τις ΗΠΑ, η πραγματικότητα είναι τελείως διαφορετική. Συγκεκριμένα οι ρίζες τους βρίσκονται στην Ευρώπη και πιο συγκεκριμένα στο Ηνωμένο Βασίλειο και την Σουηδία. Σύμφωνα με τον E.F.S.I (European Fund for Strategic Investment- Οργανισμός Επενδύσεων Ευρωπαϊκής Ένωσης) το πρώτο αμοιβαίο κεφάλαιο ιδρύθηκε στην Ελβετία το 1849 με όνομα «Societe civile Genevoise d'emploi de fonds» και ακολούθησε το 1868 στην Αγγλία με έδρα το Λονδίνο το «Foreign and Colonial Government Trust». Στις ΗΠΑ το πρώτο αμοιβαίο κεφάλαιο ιδρύθηκε το 1924 στην Μασαχουσέτη με όνομα «Massachusetts Investor Trust».

Στην Ελλάδα η εισαγωγή του θεσμού των αμοιβαίων κεφαλαίων πραγματοποιήθηκε το 1970 με το Ν.Δ 608/70 [1]. Στα τέλη του 1972 δημιουργείται από την Εμπορική Τράπεζα η πρώτη εγχώρια εταιρεία διαχείρισης με όνομα «Ελληνική» που προσφέρει το πρώτο μικτό αμοιβαίο κεφάλαιο στο κοινό με την ονομασία «Ερμής Δυναμικό» [1]. Στη συνέχεια, στις αρχές του 1973, η Εθνική Τράπεζα δημιουργεί με τη σειρά της την εταιρεία διαχείρισης «Διεθνική» που προσφέρει το αμοιβαίο κεφάλαιο «Δήλος Μικτό» [1]. Τα δύο παραπάνω αμοιβαία κεφάλαια κατείχαν όλο το μερίδιο της αγοράς μέχρι το 1989, καθώς λόγω πολιτικών

και οικονομικών γεγονότων στην ελληνική οικονομία δεν υπήρχε ιδιαίτερο ενδιαφέρον για την δημιουργία νέων αμοιβαίων κεφαλαίων.

Το 1987 πραγματοποιήθηκαν στην ελληνική κεφαλαιαγορά σημαντικές θεσμικές αλλαγές, με αποτέλεσμα να αναθερμανθεί το ενδιαφέρον για την αγορά αμοιβαίων κεφαλαίων. Έτσι το 1989 ιδρύθηκε ένα νέο αμοιβαίο κεφάλαιο από την Interamerican [1] και ακολούθησαν το 1990 άλλα τέσσερα νέα από την Alpha, την Ευρωπαϊκή Πίστη και την Ασπίς Πρόνοια.

Στην Ελλάδα όπως φαίνεται παρακάτω στο σχήμα 1.1, σύμφωνα με στοιχεία της Ένωσης Θεσμικών Επενδυτών<sup>1</sup>, υπάρχει αλματώδη ανάπτυξη της αγοράς την τελευταία δεκαετία με 14 εταιρείες διαχείρισης αμοιβαίων κεφαλαίων και συνολικά 354 διαθέσιμα αμοιβαία κεφάλαια συνολικού ενεργητικού λίγο περισσότερο από 11 δισεκατομμύρια Ευρώ.

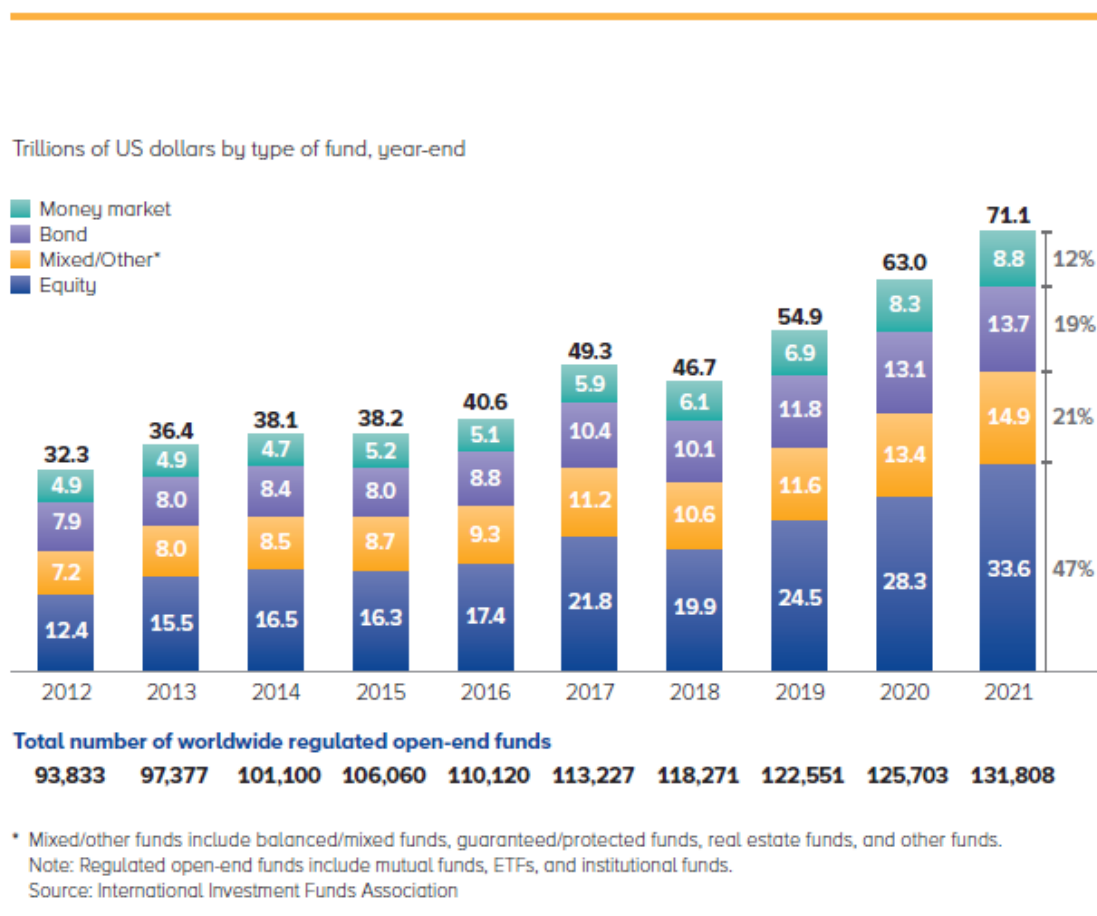
A/A	Α.Ε.Δ.Α.Κ.	Κατ. ΟΣΕΚΑ	Συν. ΟΣΕΚΑ	Συνολικό Ενεργητικό σε € την 1/1/2023	Συνολικό Ενεργητικό σε € την 19/01/2023	Δ% από 1/1/2023	Μερίδιο Αγοράς 1/1/2023	Μερίδιο Αγοράς 19/01/2023	Μεταβολή Μεριδίου Αγοράς
1	EUROBANK ASSET MANAGEMENT Α.Ε.Δ.Α.Κ.	128	128	3.078.350.838,79	3.129.701.995,31	1,67%	28,29%	27,81%	-0,48
2	ALPHA ASSET MANAGEMENT Α.Ε.Δ.Α.Κ.	50	50	2.311.927.902,76	2.475.078.485,20	7,06%	21,25%	21,99%	0,74
3	ΠΕΙΡΑΙΩΣ ASSET MANAGEMENT Α.Ε.Δ.Α.Κ.	52	52	2.042.652.542,47	2.089.259.690,48	2,28%	18,77%	18,57%	-0,20
4	Α.Ε.Δ.Α.Κ. ΑΣΦΑΛΙΣΤΙΚΩΝ ΟΡΓΑΝΙΣΜΩΝ	2	2	1.438.975.314,79	1.482.396.771,19	3,02%	13,22%	13,17%	-0,05
5	ΕΘΝΙΚΗ ASSET MANAGEMENT Α.Ε.Δ.Α.Κ.	26	26	963.482.058,23	996.894.867,71	3,47%	8,85%	8,86%	0,01
6	ALPHA TRUST	17	17	299.749.580,57	308.187.825,69	2,82%	2,75%	2,74%	-0,01
7	3K INVESTMENT PARTNERS	12	12	205.683.329,72	213.610.943,42	3,85%	1,89%	1,90%	0,01
8	TRITON ASSET MANAGEMENT Α.Ε.Δ.Α.Κ.	6	6	158.610.326,92	162.866.514,09	2,68%	1,46%	1,45%	-0,01
9	NN Α.Ε.Δ.Α.Κ.	12	12	117.904.996,92	120.695.668,23	2,37%	1,08%	1,07%	-0,01
10	ALLIANZ Α.Ε.Δ.Α.Κ.	7	7	82.242.352,60	85.088.551,42	3,46%	0,76%	0,76%	0,00
11	ΕΥΡΩΠΑΪΚΗ ΠΙΣΤΗ ASSET MANAGEMENT Α.Ε.Δ.Α.Κ.	10	10	66.556.488,87	69.047.409,11	3,74%	0,61%	0,61%	0,00
12	Optima asset management Α.Ε.Δ.Α.Κ.	7	7	55.172.289,83	56.655.760,14	2,69%	0,51%	0,50%	-0,01
13	ATTICA WEALTH MANAGEMENT Α.Ε.Δ.Α.Κ.	6	6	36.294.850,49	37.634.440,41	3,69%	0,33%	0,33%	0,00
14	ATHOS ASSET MANAGEMENT Α.Ε.Δ.Α.Κ.	19	19	24.601.476,21	26.089.039,66	6,05%	0,23%	0,23%	0,00
	<b>ΣΥΝΟΛΑ</b>	<b>354</b>	<b>354</b>	<b>10.882.204.349,17</b>	<b>11.253.207.962,06</b>	<b>3,41%</b>	<b>100,00%</b>	<b>100,00%</b>	

Σχήμα 1.1: Συνολικό ενεργητικό αγοράς ΟΣΕΚΑ ανά εταιρεία την 19/01/2023

Οι αγορές των αμοιβαίων κεφαλαίων παρουσιάζουν μεγάλους ρυθμούς ανάπτυξης παγκοσμίως την τελευταία δεκαετία [2]. Μόνο κατά το έτος 2021 σύμφωνα με το I.I.F.A (International Investment Fund Association), η αγορά αναπτύχθηκε κατά 12,85% με 131808 διαθέσιμα αμοιβαία κεφάλαια και συνολικό ενεργητικό που ανέρχεται στα \$71,1 τρισεκατομμύρια [2]. Συγκεντρωτικά τα στοιχεία της τελευταίας δεκαετίας, δηλαδή από το τέλος του 2012 μέχρι και το τέλος του 2021 φαίνονται στο Σχήμα 1.2[2]. Αξίζει να σημειωθεί ότι η αγορά αμοιβαίων κεφαλαίων έχει αναπτυχθεί κατά περίπου 120% . Στο Σχήμα 1.2

<sup>1</sup> <https://bit.ly/3IFSrST>

φαίνεται επίσης το ποσοστό των πιο δημοφιλών επενδυμένων αμοιβαίων κεφαλαίων και το μερίδιο της αγοράς που κατέχουν κατ' έτος [2].

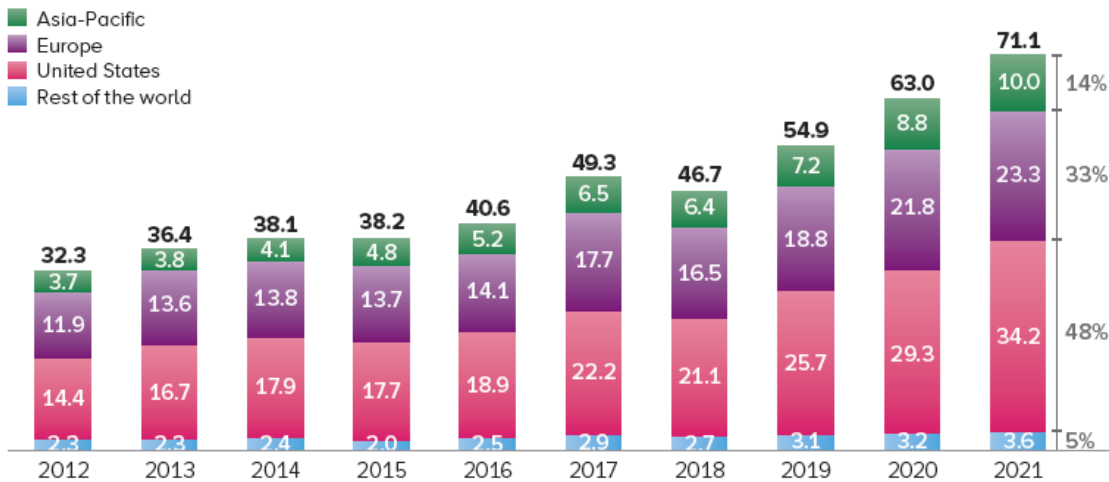


Σχήμα 1.2: Συνολικό ενεργητικό «αμοιβαίων κεφαλαίων ανοιχτού τύπου» 2012-2021

Αξίζει να σημειωθεί ότι το συνολικό ενεργητικό της αγοράς αμοιβαίων κεφαλαίων ποικίλει ανάλογα με την γεωγραφική περιοχή. Στα τέλη του 2021 οι ΗΠΑ και η Ευρώπη σύμφωνα με το I.I.F.A κατέχουν το συντριπτικά μεγαλύτερο μερίδιο της αγοράς με 48% (\$34,2 τρις) και 33% (\$23,3 τρις) αντίστοιχα [2]. Στο Σχήμα 1.3 παρουσιάζονται συγκεντρωτικά τα μερίδια του ενεργητικού με γεωγραφική κατανομή τόσο σε τρισεκατομμύρια δολάρια όσο και ποσοστιαία [2].



Trillions of US dollars by region, year-end



Note: Regulated open-end funds include mutual funds, ETFs, and institutional funds.  
Source: International Investment Funds Association

Σχήμα 1.3: Μέγεθος Αγοράς «αμοιβαίων κεφαλαίων ανοιχτού τύπου» παγκοσμίως σε τρισεκατομμύρια \$ και ποσοστιαία

## 2.2 Τι είναι αμοιβαίο κεφάλαιο

Αμοιβαίο κεφάλαιο καλείται το σύνολο της περιουσίας που σχηματίζεται από εισφορές αποταμιευτών είτε σε μετρητά είτε σε χρεόγραφα διάφορων τύπων που είναι διαπραγματεύσιμα σε μια αγορά. Το σύνολο της περιουσίας αυτής ανήκει εξ' αδιαίρετου στους μεριδιούχους, αναλογικά με τον αριθμό των μεριδίων που έχει ο καθένας στην κατοχή του<sup>2</sup>.

Στην ουσία ένα αμοιβαίο κεφάλαιο είναι ένα χρηματοπιστωτικό εργαλείο που «αντλεί» περιουσιακά στοιχεία από τους μετόχους και τα επενδύει σε μετοχές, ομόλογα, και άλλα περιουσιακά στοιχεία. Οι διαχειριστές των αμοιβαίων κεφαλαίων αποφασίζουν για το τρόπο με τον οποίο επενδύονται τα κεφάλαια, επιδιώκοντας στην παραγωγή κερδών ή εισοδήματος για τους επενδυτές του κεφαλαίου. Το χαρτοφυλάκιο ενός αμοιβαίου κεφαλαίου είναι δομημένο και κατανομημένο σύμφωνα με τους επενδυτικούς στόχους του εκάστοτε κεφαλαίου όπως αυτοί αναφέρονται στον κανονισμό του.

Τα αμοιβαία κεφάλαια παρέχουν τη δυνατότητα σε μικρούς και ανεξάρτητους επενδυτές να έχουν πρόσβαση σε χαρτοφυλάκια με επαγγελματική διαχείριση αποτελούμενα από

<sup>2</sup>Hayes A. (2022), « Mutual Funds: Different Types and How They Are Priced», Investopedia

συνδυασμό επενδύσεων. Τα αμοιβαία κεφάλαια επενδύουν σε ένα μεγάλο αριθμό χρεογράφων και η απόδοση υπολογίζεται συνήθως ως τη μεταβολή της συνολικής κεφαλαιοποίησης του αμοιβαίου κεφαλαίου, η οποία προκύπτει από τη συνολική απόδοση των υποκείμενων επενδύσεων.

Τα περισσότερα αμοιβαία κεφάλαια αποτελούν μέρος ενός μεγαλύτερου επενδυτικού ομίλου. Στο Σχήμα 1.4, για παράδειγμα, ο όμιλος 3K Investment Partners στις αρχές του 2022 διέθετε στην αγορά 10 διαφορετικά αμοιβαία κεφάλαια, με διαφορετικού τύπου επενδύσεις στο καθένα και στις 3/1/2022 εισήγαγε δύο νέα αμοιβαία κεφάλαια<sup>3</sup>.

3K INVESTMENT PARTNERS			
		Ενεργητικό σε €	Μερίδια
1	3K A/K Greek Value Μετοχικό Εσωτερικού	13.338.152,63	1.184.275,00
2	3K A/K International Income Ομολογιακό Θεσμικών	377.580,60	36.969,26
3	3K A/K International Income Ομολογιακό	7.219.031,81	1.245.449,85
4	3K A/K Μετοχικό Εσωτερικού	41.683.261,20	4.458.055,07
5	3K Μετοχικό Εσωτερικού – Κατηγορία Μεριδίων (Θ)	1.403.284,88	86.964,29
6	3K A/K Μικτό	5.720.050,95	1.141.389,24
7	NN HELLAS A/K Μετοχικό - Κατηγορία Μεριδίων (Θ)	67.618.994,43	5.285.541,09
8	NN HELLAS A/K Ομολογιακό	510.850,15	63.350,95
9	NN HELLAS A/K Ομολογιακό - Κατηγορία Μεριδίων (Θ)	28.681.050,37	1.743.053,26
10	Κατηγορία μεριδίων Θεσμικών Επενδυτών (Θ) - 3K Αμοιβαίο Κεφάλαιο Greek Value Μετοχικό Εσωτερικού	3.879.850,33	182.218,17

Σχόλια		
Ημερομηνία	ΟΣΕΚΑ	Σχόλιο
03/01/2022	3K A/K Μετοχικό Διεθνές	Έναρξη Αμοιβαίου Κεφαλαίου
03/01/2022	3K A/K Μετοχικό Διεθνές Θεσμικών Επενδυτών	Έναρξη νέας κατηγορίας μεριδίων του 3K A/K Μετοχικό Διεθνές

Σχήμα 1.4: 3K INVESTMENT PARTNERS αμοιβαία κεφάλαια- Στοιχεία ΟΣΕΚΑ  
01/01/2022

Το κάθε αμοιβαίο κεφάλαιο έχει έναν διευθυντή που πολλές φορές αναφέρεται ως «επενδυτικός σύμβουλος», ο οποίος έχει τη νομική υποχρέωση να εργάζεται και να δρα προς το συμφέρον του κεφαλαίου και των μετόχων του.

Τα αμοιβαία κεφάλαια στερούνται νομικής προσωπικότητας και το κεφάλαιο του είναι μεταβλητό (open-end fund). Η υποχρεωτική εξαγορά των μεριδίων του αμοιβαίου κεφαλαίου καθώς και η ελεύθερη (συνήθως) είσοδος των μεριδιούχων επιβεβαιώνουν την μεταβλητότητα του ενεργητικού τους. Επιπλέον κάθε αμοιβαίο κεφάλαιο χαρακτηρίζεται από πλήρη αυτονομία και ανεξαρτησία, ανεξαρτήτως του αν η εταιρεία που το διαχειρίζεται έχει στην κατοχή της άλλα αμοιβαία κεφάλαια. Η αυτονομία αυτή οφείλεται στο ότι οι αποφάσεις που λαμβάνονται αφορούν αποκλειστικά και μόνο το ενεργητικό του εκάστοτε αμοιβαίου κεφαλαίου, δηλαδή το κάθε αμοιβαίο κεφάλαιο εφαρμόζει δική του επενδυτική στρατηγική. Ο βασικός παράγοντας που προσδιορίζει τον κίνδυνο ενός αμοιβαίου κεφαλαίου έγκειται στον επενδυτικό του χαρακτήρα [2].

<sup>3</sup> <https://www.3kip.gr/el/amoivaia-kefalaia>

## 2.3 Κατηγορίες αμοιβαίων κεφαλαίων

Τα αμοιβαία κεφάλαια χωρίζονται σε κατηγορίες ανάλογα με τις επενδυτικές στρατηγικές που εφαρμόζονται από αυτά ή και την περιοχή τοποθέτησης των κεφαλαίων τους. Υπάρχουν πολλές κατηγορίες αμοιβαίων κεφαλαίων, στην παρούσα εργασία όμως θα αναφέρονται οι 10 σημαντικότερες [1], [2]:

1. Μετοχικά αμοιβαία κεφάλαια (Stock Market Mutual Funds/ Equity Mutual Funds)
2. Αμοιβαία κεφάλαια Χρηματαγοράς ή Διαχείρισης Διαθεσίμων (Money Market Mutual Funds)
3. Αμοιβαία κεφάλαια Σταθερού Εισοδήματος ( Fixed Income Mutual Funds)
4. Ομολογιακά αμοιβαία κεφάλαια (Bond Mutual Funds)
5. Μικτά αμοιβαία κεφάλαια (Mixed Mutual Funds)
6. Αμοιβαία κεφάλαια Δεικτών (Index Mutual Funds)
7. Αμοιβαία κεφάλαια Ειδικών Κατηγοριών (Specialty Mutual Funds)
8. Αμοιβαία κεφάλαια Εμπορευμάτων (Commodities Mutual Funds)
9. Διεθνή αμοιβαία κεφάλαια (International/ Global Mutual Funds)
10. ETFs (Exchange Traded Funds)

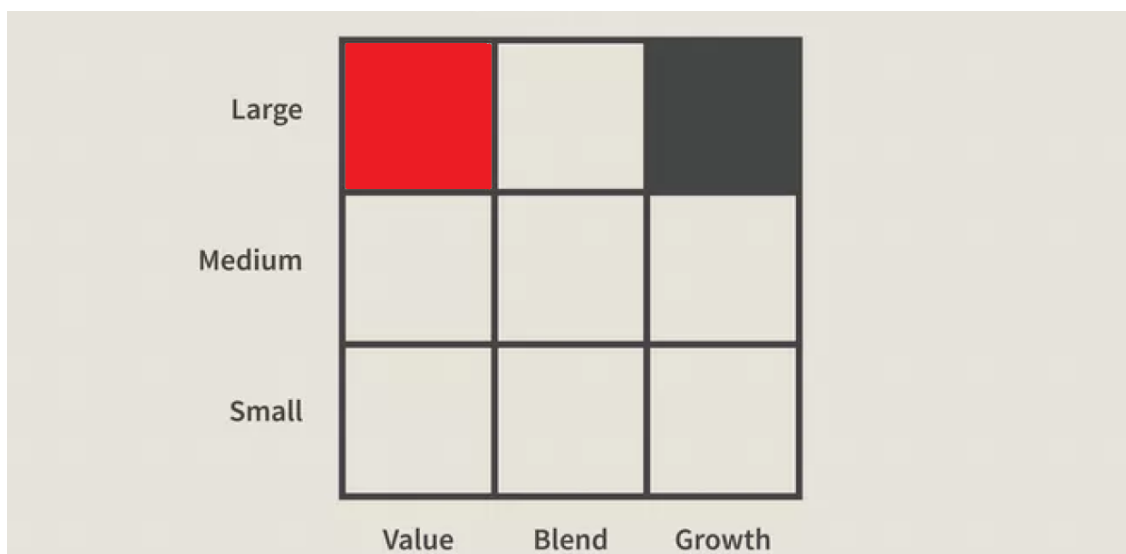
Παρακάτω θα γίνει μία συνοπτική ανάλυση της κάθε κατηγορίας:

### Μετοχικά αμοιβαία κεφάλαια (Stock Market Mutual Funds/ Equity Mutual Funds)

Όπως υποδηλώνει η ονομασία τους, τα μετοχικά αμοιβαία κεφάλαια επενδύουν κυρίως σε μετοχές <sup>1,2</sup> [1], [2]. Σε αυτήν την κατηγορία υπάρχουν πολλές υποκατηγορίες κεφαλαίων. Σε κάποια μετοχικά αμοιβαία κεφάλαια το όνομά τους υποδηλώνει τον τύπο των εταιρειών στις οποίες επενδύουν (small-, mid- ή large-cap). Οι large-cap εταιρείες έχουν υψηλή κεφαλαιοποίηση, οι mid-cap μέση κεφαλαιοποίηση και οι small-cap μικρή κεφαλαιοποίηση. Σε άλλα μετοχικά αμοιβαία κεφάλαια το όνομά τους υποδηλώνει τον τύπο της επενδυτικής προσέγγισης τους (aggressive growth, income- oriented, value κ.α.). Επίσης τα μετοχικά αμοιβαία κεφάλαια χωρίζονται σε υποκατηγορίες ανάλογα με το αν επενδύουν σε εγχώριες ή ξένες μετοχές (εσωτερικού ή εξωτερικού). Η ονομασία ενός μετοχικού αμοιβαίου κεφαλαίου μπορεί να συνδυάζει και τους τρεις παραπάνω προσδιοριστικούς παράγοντες (πχ Large-Cap Oriented Domestic Fund).

Οι στρατηγικές που ακολουθούνται από τα μετοχικά αμοιβαία κεφάλαια μπορούν να αποτελούνται από ένα μείγμα μεταξύ του στυλ επένδυσης και του μεγέθους των εταιρειών

που επενδύουν. Στο Σχήμα 1.5 παρουσιάζεται ένας τρόπος προσδιορισμού των επενδύσεων για τα Μετοχικά αμοιβαία κεφάλαια.



Σχήμα 1.5: Προσδιορισμός επενδύσεων με βάση την κεφαλαιοποίηση των εταιρειών και το είδος των επενδύσεων στις οποίες επενδύουν τα μετοχικά αμοιβαία κεφάλαια (Πηγή:[4])

Για παράδειγμα, ένα large-cap value fund θα επενδύσει σε εταιρείες υψηλής κεφαλαιοποίησης με πρόσφατη πτώση στην τιμή των μετοχών τους. Το στυλ της επένδυσης αυτής τοποθετείται στο πάνω αριστερό τετράγωνο του σχήματος (κόκκινο).

Γενικά θα μπορούσαμε να πούμε ότι τα Μετοχικά αμοιβαία κεφάλαια, θεωρούνται επενδύσεις υψηλού κινδύνου, καθώς για το συγκεκριμένο περιουσιακό στοιχείο δεν παρέχεται καμία εγγύηση ή ασφάλεια για το επίπεδο της απόδοσής του. Αυτό θα μπορούσαμε να πούμε ότι οφείλεται στο γεγονός ότι οι μετοχές δεν προσφέρουν σταθερό ή βέβαιο μέρισμα εκτός από ένα ελάχιστο υποχρεωτικό υπό την προϋπόθεση ότι η μετοχή έχει καταγράψει κέρδη.

#### Αμοιβαία κεφάλαια Χρηματαγοράς ή Διαχείρισης Διαθεσίμων (Money Market Mutual Funds)

Τα αμοιβαία κεφάλαια της εν λόγω κατηγορίας επενδύουν σε χρεόγραφα πιο ασφαλή από τα μετοχικά κεφάλαια, ενώ ταυτόχρονα έχουν μικρή χρονική διάρκεια λήξης [1]. Τα «Money Market Mutual Funds» υποχρεώνονται εξ' ορισμού σε πιο ασφαλείς επενδύσεις, με έμφαση σε επενδύσεις με σχεδόν μηδενικό κίνδυνο όπως τα κρατικά ομόλογα ή επιλεγμένα ομόλογα μεγάλων εταιρειών χαμηλού κινδύνου. Γενικά παρέχουν σταθερή (χαμηλή) απόδοση, η οποία προσδιορίζεται από τα ισχύοντα επιτόκια της αγοράς, επενδύοντας κυρίως σε βραχυπρόθεσμα

χρεόγραφα που είναι διαθέσιμα στη χρηματαγορά (πχ., έντοκα γραμμάτια δημοσίου, προθεσμιακές καταθέσεις, εμπορικά γραμμάτια κ.ά.).

### Αμοιβαία Κεφάλαια Σταθερού Εισοδήματος ( Fixed Income Mutual Funds)

Τα αμοιβαία κεφάλαια της εν λόγω κατηγορίας έχουν κύριο στόχο την παροχή στο μέτρο του δυνατού, «σταθερού εισοδήματος» στους επενδυτές [1], [2]. Επενδύουν κυρίως σε τίτλους οι οποίοι προσφέρουν όσο το δυνατόν γίνεται πιο «βέβαιη» απόδοση ετησίως ή σε τακτικά χρονικά διαστήματα (εξάμηνα-τρίμηνα ή μηνιαία). Τέτοιοι τίτλοι είναι τα ομόλογα κρατικών εταιρειών, τα ομόλογα δημοσίου, ομόλογα τραπεζών καθώς και τα ομόλογα επιλεγμένων εταιρειών με υψηλή πιστοληπτική ικανότητα. Ο κύριος στόχος αυτών των αμοιβαίων κεφαλαίων είναι να εξασφαλίσουν μία σταθερή ροή εσόδων για τους επενδυτές, είτε αυτή πραγματοποιείται υπό την μορφή συγκεκριμένου τόκου είτε υπό μορφή μερίσματος που καταβάλλεται από την εκδότρια εταιρεία στο αμοιβαίο κεφάλαιο.

### Ομολογιακά αμοιβαία κεφάλαια (Bond Mutual Funds)

Τα αμοιβαία κεφάλαια ομολογιών επενδύουν κυρίως σε πιο μεσοπρόθεσμες ή/και μακροπρόθεσμες ομολογίες που εκδίδονται είτε από κράτη είτε από μεγάλες εταιρείες είτε ακόμη και από δήμους (στις ΗΠΑ οι δήμοι εκδίδουν δημοτικά ομόλογα- Municipal Bonds)<sup>2</sup> [1]. Γενικά, ο κίνδυνος σε σχέση με τα κεφάλαια σταθερού εισοδήματος είναι αισθητά μεγαλύτερος, γεγονός που οφείλεται στη διάρκεια των επενδύσεων. Συνήθως όμως λόγω της παροχής σημαντικού εισοδήματος κατά τη διάρκεια της κατοχής ενός ομολόγου μέχρι και την ωρίμανσή του (από τόκους, τοκομερίδια, κλπ.), ο κίνδυνος αυτός αντισταθμίζεται με τη μεγαλύτερη απόδοση που προσφέρουν.

### Μικτά αμοιβαία κεφάλαια (Mixed Mutual Funds)

Τα μικτά αμοιβαία κεφάλαια επενδύουν σε ένα σύνολο προνομιούχων μετοχών και ομολόγων, έχοντας ως στόχο την επίτευξη όσο το δυνατόν καλύτερης διαφοροποίησης που περιορίζει τον κίνδυνο σε ανεκτά για τους επενδυτές επίπεδα [1], [2]. Γενικά απευθύνονται σε συντηρητικούς επενδυτές και ο κύριος επενδυτικός σκοπός τους είναι η διατήρηση του αρχικού κεφαλαίου των επενδυτών (μακροπρόθεσμα), παρέχοντας τους παράλληλα ένα τακτικό μέρισμα για την κάλυψη των τρεχουσών αναγκών τους.

## Αμοιβαία κεφάλαια Δεικτών (Index Mutual Funds)

Τα αμοιβαία κεφάλαια δεικτών έχουν σαν στόχο την επένδυση σε εμπορεύματα, μετοχές ή ομολογίες των οποίων η απόδοση ακολουθεί την απόδοση κάποιων συγκεκριμένων χρηματιστηριακών δεικτών (πχ., S&P 500, Dow Jones Industrial Average DJIA, κ.ά.) [1],[2]. Οι υπάρχοντες δείκτες απαρτίζονται από εταιρείες συγκεκριμένου κλάδου της οικονομίας (πχ., βιομηχανία, υψηλές τεχνολογίες, ανανεώσιμες πηγές ενέργειας) ή από συγκεκριμένα εμπορεύματα (χρυσό, σιτηρά, λάδι). Τα αμοιβαία κεφάλαια δεικτών επενδύουν στους τίτλους που απαρτίζουν τον δείκτη που έχουν επιλέξει κατά ποσοστιαία αναλογία των τίτλων που συμμετέχουν σε αυτόν. Οποιαδήποτε αλλαγή τίτλων στο ποσοστό συμμετοχής στον δείκτη πρέπει να οδηγήσει και σε αλλαγή στην δομή της επένδυσης από το αμοιβαίο κεφάλαιο. Η εν λόγω στρατηγική προϋποθέτει λιγότερη έρευνα από τους χρηματοοικονομικούς αναλυτές και συμβούλους, με αποτέλεσμα να περιορίζονται τα έξοδα που περνούν στους μετόχους (για παροχή συμβουλευτικών ή/και διοικητικών υπηρεσιών). Με τον τρόπο αυτό η απόδοση του αμοιβαίου κεφαλαίου είναι ανάλογη της απόδοσης του δείκτη που έχει επιλέξει να στηριχτεί για τη διαμόρφωση του χαρτοφυλακίου επενδύσεών του.

## Αμοιβαία κεφάλαια ειδικών κατηγοριών (Specialty Mutual Funds)

Στην κατηγορία αυτή ανήκουν τα αμοιβαία κεφάλαια που οι επενδύσεις τους επικεντρώνονται αποκλειστικά σε συγκεκριμένους κλάδους της οικονομίας ή και σε συγκεκριμένες περιοχές [1], [2]. Στην κατηγορία αυτή έχουμε 3 μεγάλες υποκατηγορίες :

- Κλαδικά κεφάλαια (Sector Funds) που επενδύουν σε εταιρείες που ανήκουν σε συγκεκριμένους επιχειρηματικούς κλάδους (βιομηχανία, νέες τεχνολογίες, υγεία, ενέργεια, κ.ά.). Τα κλαδικά αμοιβαία κεφάλαια χαρακτηρίζονται από πολύ υψηλό κίνδυνο ο οποίος προέρχεται κυρίως από τις διακυμάνσεις της οικονομικής κατάστασης στον συγκεκριμένο κλάδο. Αξίζει να σημειωθεί ότι όλοι οι κλάδοι της οικονομίας συσχετίζονται μεταξύ τους, άρα ο κίνδυνος των κλαδικών κεφαλαίων προσανυξάνεται από διακυμάνσεις και σε άλλους κλάδους της οικονομίας.
- Regional Funds: είναι τα αμοιβαία κεφάλαια που επιλέγουν να επενδύσουν σε εταιρείες μίας συγκεκριμένης γεωγραφικής περιοχής. Η εστίαση στην περιοχή μπορεί να είναι σε επίπεδο ηπείρων, κρατών ή/ και ακόμη και περιφερειών.
- Responsible Funds: άλλη μία υποκατηγορία των αμοιβαίων κεφαλαίων ειδικών κατηγοριών, όπου επιλέγουν να μην επενδύουν τα κεφάλαιά τους σε κλάδους που εντάσσονται στο πλαίσιο «κοινωνικά μη υπεύθυνων» επενδύσεων, όπως καπνοβιομηχανίες, παραγωγή αλκοολούχων ποτών, βιομηχανία όπλων ή πυρηνικής

ενέργειας, κ.ά. Συνήθως τα responsible funds επενδύουν κυρίως σε κλάδους πιο φιλικούς στο περιβάλλον όπως ηλιακά πάρκα, στην αιολική ενέργεια, στον τομέα της ανακύκλωσης, κ.ά.

### Αμοιβαία κεφάλαια εμπορευμάτων (Commodities Mutual Funds)

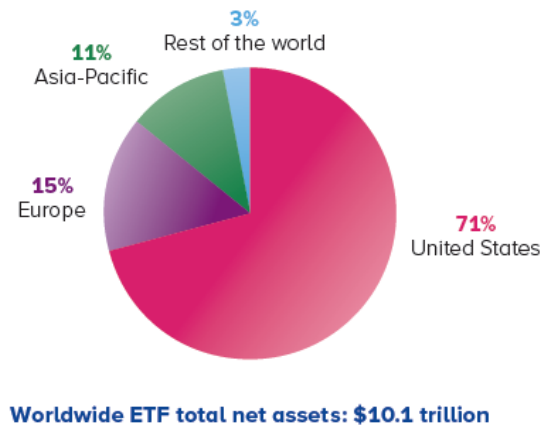
Τα αμοιβαία κεφάλαια της συγκεκριμένης κατηγορίας επενδύουν τα κεφάλαιά τους σε διάφορα προϊόντα όπως πολύτιμα μέταλλα, καθώς και σε προθεσμιακά συμβόλαια εμπορευμάτων και χρηματοπιστωτικών τίτλων των οποίων οι αγοραπωλησίες είναι διαπραγματεύσιμες στις αγορές [1], [2]. Γενικά χαρακτηρίζονται από μεγάλη ρευστότητα αλλά και από υψηλό κίνδυνο παρά τη διαφοροποίηση που επιτυγχάνεται.

### Διεθνή ή διεθνικά αμοιβαία κεφάλαια (International/ Global Mutual Funds)

Τα διεθνή αμοιβαία κεφάλαια επενδύουν μπορεί να επενδύσουν τα κεφάλαιά τους σε χρεόγραφα τα οποία εκδίδονται και διαπραγματεύονται σε όλον τον κόσμο [1], [2]. Κατά συνέπεια, τα διεθνή αμοιβαία κεφάλαια ενσωματώνουν τόσο τον πολιτικό κίνδυνο (της περιοχής στην οποία επενδύουν) όσο και τον συναλλαγματικό κίνδυνο. Ο συναλλαγματικός κίνδυνος μπορεί να μετριαστεί, παρέχοντας ευκαιρίες συμμετοχής στην άνοδο που προσφέρουν κάποιες άλλες αγορές χρήματος και κεφαλαίου. Η επένδυση σε Διεθνή αμοιβαία κεφάλαια μπορεί να αποτελεί μέρος ενός ισορροπημένου χαρτοφυλακίου, αυξάνοντας το επίπεδο διαφοροποίησης και μειώνοντας τον συνολικό κίνδυνο, καθώς οι αποδόσεις σε ξένες χώρες μπορεί να μην συσχετίζονται άμεσα ή έμμεσα με της αποδόσεις των εγχώριων εταιρειών.

### ETFs (Exchange Traded Funds)

Ένα ETF είναι μία εγγεγραμμένη επενδυτική εταιρεία που μοιάζει με τα αμοιβαία κεφάλαια αλλά δεν κατατάσσεται στην ίδια κατηγορία [2]. Επενδύουν τα κεφάλαιά τους σε μετοχές, ομόλογα και άλλα περιουσιακά στοιχεία όπως παράγωγα ή προθεσμιακές καταθέσεις. Εφαρμόζουν τις ίδιες επενδυτικές στρατηγικές με τα αμοιβαία κεφάλαια με τη βασική τους διαφορά ότι μπορούν να αγοραστούν/ πουληθούν και κατά τη διάρκεια της ημέρας. Ένα ETF τυπικά έχει μικρότερες χρεώσεις από ένα αντίστοιχο αμοιβαίο κεφάλαιο. Γενικά η αγορά των ETFs γνωρίζει ιδιαίτερη ανάπτυξη στις ΗΠΑ, λιγότερο στην Ευρώπη, και είναι ακόμη σε πολύ χαμηλά επίπεδα στον υπόλοιπο κόσμο, όπως φαίνεται και στο Σχήμα 1.6.



Sources: Investment Company Institute and ETFGI

---

Σχήμα 1.6: Ποσοστό δραστηριοποίησης ETFs σε κάθε ήπειρο σε σχέση με την παγκόσμια αγορά (Πηγή: [3])

Όπως φαίνεται στο Σχήμα 1.6 το μέγεθος της αγοράς των ETFs στα τέλη του 2021 ανέρχεται στα \$10,1 τρισεκατομμύρια με το 71% να δραστηριοποιείται στις ΗΠΑ, το 15% στην Ευρώπη και το υπόλοιπο 14% στον υπόλοιπο κόσμο.

## 2.4 Πλεονεκτήματα-μειονεκτήματα αμοιβαίων κεφαλαίων

Στην ενότητα αυτή θα αναφερθούν συνοπτικά τα πλεονεκτήματα και τα μειονεκτήματα της επένδυσης σε εταιρείες αμοιβαίων κεφαλαίων όπως καταγράφονται στην βιβλιογραφία [1], [2].

### Πλεονεκτήματα της επένδυσης σε αμοιβαία κεφάλαια

Τα σημαντικότερα πλεονεκτήματα που προσφέρει στους μεριδιούχους η επένδυση σε αμοιβαία κεφάλαια είναι τα παρακάτω:

- Έμπειρη και συνεχής επαγγελματική διαχείριση των κεφαλαίων από επαγγελματίες του κλάδου, εξασφαλίζοντας ορθολογική χρήση των κεφαλαίων.
- Μεγάλη ρευστότητα



- Ποικιλία επιλογών για τους επενδυτές καθώς υπάρχουν πολλές διαφορετικές κατηγορίες, σε σχεδόν όλους τους τομείς της οικονομίας.
- Μείωση του επενδυτικού κινδύνου μέσω της διαφοροποίησης. Μελέτες έχουν δείξει ότι μέσω της διαφοροποίησης (δηλαδή τη διασπορά του κινδύνου μέσω πολλών διαφορετικών επενδύσεων στο χαρτοφυλάκιο) επιτυγχάνεται σε μεγάλο βαθμό η προστασία της περιουσίας των επενδυτών καθώς και η επίτευξη σχετικά σταθερών κερδών.
- Διαφάνεια στις συναλλαγές και απλουστευμένες διαδικασίες στην παρακολούθηση της εξέλιξης μιας επένδυσης
- Παρέχεται η δυνατότητα επανεπένδυσης των μερισμάτων δίχως την καταβολή προμήθειας
- Πληρωμές μερισμάτων από τα κέρδη: Πολλά αμοιβαία κεφάλαια ιδιαίτερα τα αμοιβαία κεφάλαια Σταθερού Εισοδήματος (Fixed Income Mutual Funds) εφαρμόζουν πληρωμές μερισμάτων στους μεριδιούχους. Όταν ένα κεφάλαιο έχει κέρδη τακτικά μπορεί να μοιράζει ένα ποσοστό των κερδών στους μετόχους (αναλογικά).

### Μειονεκτήματα της επένδυσης σε αμοιβαία κεφάλαια

Τα σημαντικότερα μειονεκτήματα από την επένδυση σε αμοιβαία κεφάλαια είναι τα παρακάτω:

- Οι επενδυτές δεν μπορούν να επιλέξουν οι ίδιοι τις επενδύσεις στις οποίες θα κατευθυνθούν τα κεφάλαιά τους. Πολλοί θεωρούν την επαγγελματική διαχείριση συχνά μη αποτελεσματική ή και ιδιοτελή, γι' αυτό και προτιμούν να επιλέγουν οι ίδιοι που θα επενδύσουν. Για αυτούς τους επενδυτές δεν συνιστάται η επιλογή της επένδυσης σε αμοιβαία κεφάλαια.
- Γενικά τα αμοιβαία κεφάλαια θεωρούνται επενδυτικές επιλογές με χαμηλότερο κίνδυνο (λόγω της διαφοροποίησης) παρέχοντας όμως μικρότερα προσδοκώμενα κέρδη. Άρα για έναν ριψοκίνδυνο επενδυτή δεν αποτελούν την σωστή επενδυτική πρόταση
- Σε περιόδους έντονης αστάθειας δημιουργούνται προβλήματα στους διαχειριστές των κεφαλαίων λόγω των σπασμωδικών κινήσεων των μεριδιούχων, που ρευστοποιούν άμεσα (συχνά και μαζικά) τις μετοχές τους.
- Πολλά αμοιβαία κεφάλαια χρεώνουν υψηλές αμοιβές διάθεσης και εξαγοράς, καθώς και προμήθειες διαχείρισης, χρεώσεις ρευστοποίησης και άλλες χρεώσεις, οι οποίες μειώνουν την καθαρή απόδοση του αμοιβαίου κεφαλαίου .
- Δεν υπάρχει καμία εγγύηση του κεφαλαίου του επενδυτή, ούτε εγγύηση της απόδοσης των επενδύσεων του αμοιβαίου κεφαλαίου.
- Υπάρχει δυσκολία στην σύγκριση των όμορων αμοιβαίων κεφαλαίων

- Μπορεί κάποια αμοιβαία κεφάλαια να επιτρέπουν στον επενδυτή να αιτηθεί την μετατροπή του μεριδίου του σε ρευστό οποιαδήποτε στιγμή, αντίθετα, πολλά αμοιβαία κεφάλαια επιτρέπουν την εξαγορά ή ρευστοποίηση μεριδίων μόνο στο τέλος της επενδυτικής ημέρας.

Συνοψίζοντας λοιπόν μπορούμε να πούμε ότι:

- Τα αμοιβαία κεφάλαια είναι ένα επενδυτικό εργαλείο που το χαρτοφυλάκιο του αποτελείται από έναν συνδυασμό μετοχών, ομολόγων και άλλων αξιογράφων
- Τα αμοιβαία κεφάλαια κατατάσσονται σε συγκεκριμένου τύπου κατηγορίες ανάλογα με το είδος των αξιογράφων στα οποία επενδύουν, τους επενδυτικούς στόχους που θέτουν αλλά και το είδος των κερδών που φιλοδοξούν να έχουν.
- Τα αμοιβαία κεφάλαια χρεώνουν ετήσιες χρεώσεις, προμήθειες και διάφορα άλλα έξοδα, τα οποία επηρεάζουν την συνολική απόδοση και πρέπει να λαμβάνονται υπόψιν από τους επενδυτές.
- Γενικά υπάρχει η παγκόσμια τάση συνταξιοδοτικά πλάνα τα οποία χρηματοδοτούνται από εργοδότες να επενδύουν σε αμοιβαία κεφάλαια, μακροχρόνιας χαμηλής απόδοσης και μικρού κινδύνου (πχ., τα Ομολογιακά αμοιβαία κεφάλαια)
- Τα αμοιβαία κεφάλαια αποτελούν σημαντικό επενδυτικό εργαλείο με πολλά πλεονεκτήματα αλλά και σημαντικά μειονεκτήματα που πρέπει να λαμβάνει υπόψιν ο εκάστοτε επενδυτής.

### 3. ΜΕΘΟΔΟΙ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΗΝ ΑΞΙΟΛΟΓΗΣΗ ΑΜΟΙΒΑΙΩΝ ΚΕΦΑΛΑΙΩΝ

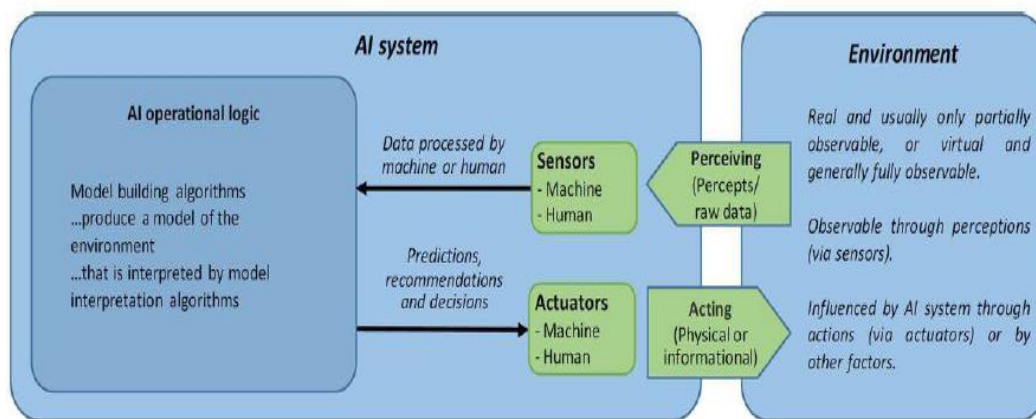
#### 3.1 Το πλαίσιο της τεχνητής νοημοσύνης

Τα συστήματα τεχνητής νοημοσύνης (Artificial Intelligence- AI) είναι συστήματα βασισμένα σε μηχανές- υπολογιστές με ποικίλα επίπεδα αυτονομίας [3]. Τα εν λόγω συστήματα μπορούν για ένα δεδομένο σύνολο δεδομένων, να κάνουν προβλέψεις, προτάσεις ή ακόμη και να λάβουν αποφάσεις χρησιμοποιώντας έναν τεράστιο όγκο δεδομένων από μία ή εναλλακτικές πηγές δεδομένων που αναφέρονται ως «big data». Τέτοιου τύπου δεδομένα τροφοδοτούν συστήματα- μοντέλα μηχανικής μάθησης (Machine Learning- ML) τα οποία έχουν την δυνατότητα να «εκπαιδεύονται» μόνα τους από ένα σύνολο δοσμένων δεδομένων και να βελτιώνονται χωρίς ανθρώπινη παρέμβαση.

Ένα σύστημα τεχνητής νοημοσύνης είναι λοιπόν ένα σύστημα που μπορεί για ένα δεδομένο σύνολο καθορισμένων δεδομένων να κάνει προβλέψεις συστάσεις ή να λάβει αποφάσεις επηρεάζοντας πραγματικά ή εικονικά περιβάλλοντα [3]. Χρησιμοποιεί μηχανικά ή ανθρώπινα δεδομένα για να αντιληφθούν πραγματικά ή εικονικά περιβάλλοντα, χρησιμοποιώντας το αποτέλεσμα του μοντέλου για την διατύπωση πληροφοριών ή προτεινόμενης δράσης. Υπάρχουν πολλά επίπεδα σχεδιασμού και ποικιλία αυτονομίας στα συστήματα τεχνητής νοημοσύνης. Ο κύκλος δράσης ενός συστήματος τεχνητής νοημοσύνης σύμφωνα με το AI Experts Group (AIGO) χωρίζεται στις παρακάτω φάσεις [3] :

- i. Προγραμματισμός και σχεδιασμός (Planning and Design), συλλογή δεδομένων, επεξεργασία, δημιουργία μοντέλου, ερμηνεία
- ii. Επαλήθευση και επικύρωση (Verification and Validation)
- iii. Ανάπτυξη (Deployment)
- iv. Λειτουργία και παρακολούθηση (Operation and Monitoring)

Στο σχήμα 2.1 αποτυπώνεται η λογική της λειτουργίας ενός συστήματος τεχνητής νοημοσύνης [5]:



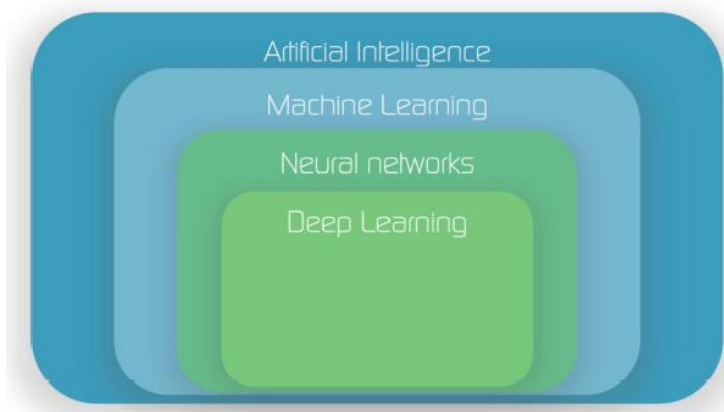
Σχήμα 2.1: Λειτουργία συστήματος τεχνητής νοημοσύνης AI Experts Group

Τα συστήματα μηχανικής μάθησης (Machine Learning- ML), που αποτελούν ένα υποσύνολο των συστημάτων τεχνητής νοημοσύνης, περιγράφουν την ικανότητα του λογισμικού να «μάθει» από τα διαθέσιμα σύνολα δεδομένων και να αυτό-βελτιώνεται χωρίς την επέμβαση του ανθρώπινου παράγοντα. Διάφοροι τύποι μηχανικής μάθησης όπως η μάθηση με επίβλεψη (supervised learning), η μάθηση χωρίς επίβλεψη (unsupervised learning), η βαθιά μάθηση (deep learning) και η ενισχυτική μάθηση (reinforcement learning).

- Η επιβλεπόμενη μάθηση (supervised learning) αποκαλείται και ως «κλασσική μηχανική μάθηση», αποτελείται από προχωρημένες παλινδρομήσεις και κατηγοριοποιήσεις των δεδομένων με στόχο να βελτιωθούν οι προβλέψεις,
- Η μη επιβλεπόμενη μάθηση (unsupervised learning) επικεντρώνεται στην επεξεργασία των δεδομένων εισόδου με σκοπό την «κατανόηση» της κατανομής των δεδομένων (πχ για τη διαμόρφωση ομάδων πελατών)
- Μια κατηγορία βαθιάς μάθησης είναι τα βαθιά νευρωνικά δίκτυα (deep learning neural networks –DLNN) τα οποία στην ουσία μιμούνται τον τρόπο που αλληλοεπιδρούν οι βιολογικοί νευρώνες στον ανθρώπινο εγκέφαλο με πολλές στρώσεις- επίπεδα (εξ' ου και ο όρος deep) προσομοιωμένων διασυνδέσεων. Η πηγή έμπνευσης αυτών των συστημάτων ήταν η λειτουργικότητα του ανθρώπινου εγκεφάλου όπου εκατοντάδες δισεκατομμύρια διασυνδεδεμένων νευρώνων μεταφέρουν πληροφορίες παράλληλα [4]. Τα νευρωνικά δίκτυα αποτελούνται από θεμελιώδης μονάδες, όπως οι ανθρώπινοι νευρώνες, που συνδέονται μεταξύ τους με δεσμούς μεταβλητούς, αποτέλεσμα της διαδικασίας της μάθησης. Τέτοια μοντέλα μπορούν να αναγνωρίσουν και να κατηγοριοποιήσουν τα δεδομένα εισόδου χωρίς να είναι απαραίτητο να του δοθούν συγκεκριμένοι κανόνες και μπορούν να αναγνωρίσουν νέες διόδους για την εύρεση λύσης που κανένας ανθρώπινος εγκέφαλος δεν θα μπορούσε να έχει αντιληφθεί ότι υπάρχουν ή ότι μπορούν να

αναπτυχθούν [5]. Αυτά τα δίκτυα θεωρούνται ότι έχουν μεγαλύτερη ανοχή στον θόρυβο (των δεδομένων) και μπορούν να λειτουργήσουν σε πολλαπλά επίπεδα.

Στο Σχήμα 2.2 φαίνεται μία παραστατική απεικόνιση του συνόλου «Τεχνητής Νοημοσύνης» και των βασικών υποσυνόλων της.<sup>4</sup>



Σχήμα 2.2: Απεικόνιση των υποσυνόλων τεχνητής νοημοσύνης (Πηγή: Hackermoon.com)

Η κρίση της νόσου Covid-19 επιτάχυνε και έκανε ακόμη πιο επιτακτική την ανάγκη για ψηφιοποίηση, μία τάση που είχε παρατηρηθεί πριν την πανδημική κρίση, γύρω δηλαδή από την χρήση συστημάτων τεχνητής νοημοσύνης. Ο τομέας των χρηματοοικονομικών (διαχείριση περιουσιακών στοιχείων, συναλλαγές μέσω αλγορίθμων, χρηματοοικονομικές συναλλαγές κρυπτονομισμάτων) δεν θα μπορούσε να εξαιρεθεί από την νέα τάση της εποχής. Η ανάπτυξη μεθόδων μηχανικής μάθησης και συστημάτων τεχνητής νοημοσύνης εξυπηρετείται από την αφθονία διαθέσιμων δεδομένων για κάθε τομέα καθώς και τη διαθέσιμη υπολογιστική ισχύ και χωρητικότητα για την επεξεργασία και αποθήκευσή τους.

Γενικά η χρήση συστημάτων τεχνητής νοημοσύνης είναι ενσωματωμένη σε προϊόντα και υπηρεσίες σε πολλούς κλάδους της βιομηχανίας (φαρμακοβιομηχανία, αυτοκινητοβιομηχανία κ.ά.) και αναπτύσσονται ολοένα και περισσότερο από παρόχους χρηματοοικονομικών υπηρεσιών σε όλους τους κλάδους του χρηματοπιστωτικού τομέα. Για παράδειγμα στον τραπεζικό τομέα (πχ αξιολόγηση επιχειρήσεων, εκτίμηση πιστωτικού κινδύνου, παρακολούθηση και εντοπισμός πιθανής απόπειρας εξαπάτησης, εξυπηρέτηση πελατών κ.ά.), στον τομέα της διαχείρισης ενεργητικού (στρατηγικές διαχείρισης χαρτοφυλακίου, διαχείριση κινδύνου κ.ά.), στον τομέα των συναλλαγών (αλγοριθμικές συναλλαγές) και στον τομέα των ασφαλειών (διαχείριση απαιτήσεων).

<sup>4</sup> Hackermoon.com (2020), «Why AI + Blockchain Make Sense », Hacker Noon, <https://hackernoon.com/why-ai-blockchain-make-sense-5k4u3s6l>

### 3.2 Αναγκαιότητα ανάπτυξης μεθόδων Μηχανικής Μάθησης

Μετά την οικονομική κρίση που ξέσπασε στις ΗΠΑ το 2008 πολλά μοντέλα πρόβλεψης που στηρίζονταν σε ποσοτικούς παράγοντες, καθώς και τα συμβατικά μοντέλα αποτύγχαναν να κατευθύνουν με επιτυχία τους επενδυτές. Για τον λόγο αυτό πολλοί επαγγελματίες ξεκίνησαν την ανάπτυξη χειροποίητων μοντέλων που μπορούσαν να μαθαίνουν από τα δεδομένα αλλά και αυτά αποδείχτηκαν χαμηλής αξιοπιστίας. Για πολλά χρόνια οι επενδυτές βασίστηκαν σε οικονομετρικές προσεγγίσεις, αλλά μόνο ελάχιστες είχαν επιτυχία χρησιμοποιώντας μοντέλα βασισμένα σε αυτές τις τεχνικές. Αυτό μπορεί να οφείλεται σε διάφορους παράγοντες όπως [6]:

- Θορύβου στα χρηματοοικονομικά δεδομένα
- Κάποιοι παράγοντες ενδέχεται να είναι πολυσυγγραμμικοί
- Η σύνδεση μεταξύ μεταβλητών και αποδόσεων μπορεί να είναι μεταβαλλόμενες, μη γραμμικές, με ή χωρίς συνάφεια μεταξύ τους.

Οι παραπάνω λόγοι κάνουν την εκτίμηση, για οποιαδήποτε διασύνδεση μεταξύ πιθανών παραγόντων πρόβλεψης, προβληματική για τα παραδοσιακά μοντέλα

Δεδομένου λοιπόν των περιορισμένων δυνατοτήτων των «χειροποίητων» μοντέλων, έγινε επιτακτική η ανάγκη ανάπτυξης τεχνικών μηχανικής μάθησης, οι οποίες μαθαίνουν αυτόματα μέσω των δεδομένων που λαμβάνουν. Οι Gu et al. [7] προσπάθησαν να δώσουν έναν ορισμό της «μηχανικής μάθησης». Σύμφωνα με τους συγγραφείς «μηχανική μάθηση είναι μία πολυποίκιλη συλλογή από πολυδιάστατα μοντέλα για στατιστική πρόβλεψη, συνδυασμένα με μεθόδους κανονικοποίησης, για την επιλογή μοντέλου και την μείωση της υπερπροσαρμογής, και αποδοτικοί αλγόριθμοι για διερεύνηση μέσα από έναν μεγάλο αριθμό πιθανών χαρακτηριστικών μοντέλων.»

Αυτού του είδους μέθοδοι έχουν μεγάλη επιτυχία σε πολλούς τομείς. Ένας σημαντικός παράγοντας επιτυχίας τους είναι ότι οι μέθοδοι αυτοί έχουν την ικανότητα γενίκευσης σε νέα δεδομένα, να μαθαίνουν από κατανομές με μεγάλο θόρυβο, και να εντοπίζουν σύνθετα χαρακτηριστικά περίπλοκων δεδομένων. Με άλλα λόγια υπερτερούν εκεί όπου οι κλασικές προσεγγίσεις αντιμετωπίζουν σημαντικούς περιορισμούς. Η ραγδαία αύξηση των εφαρμογών και των μοντέλων μηχανικής μάθησης, οφείλεται σε μία σειρά εξελίξεων στον τεχνολογικό κλάδο. Τέτοιες εξελίξεις ήταν για παράδειγμα η ανάπτυξη υπολογιστικών συστημάτων με μεγάλη ισχύ, η αύξηση διαθέσιμου χώρου αποθήκευσης δεδομένων, η διαθεσιμότητα μεγάλου όγκου δεδομένων, η ανάπτυξη νέων ισχυρών τεχνικών βελτιστοποίησης κ.ά.

Κάποια παραδείγματα εφαρμογών μηχανικής μάθησης είναι:

- Εφαρμογές πρόβλεψης μέσω χρονοσειρών
- Εφαρμογές τμηματοποίησης της αγοράς
- Διαχείριση περιουσιακών στοιχείων (asset management) κ.ά.

Αυτό που πρέπει να λαμβάνουμε υπόψιν είναι ότι οι χρηματοοικονομικές αγορές έχουν ένα σημαντικό χαρακτηριστικό [8]. Αντανακλούν τις πράξεις των ανθρώπων οι οποίες μπορεί να έχουν επηρεαστεί από άλλους ή από προηγούμενες έρευνες που έχουν γίνει. Οι μέθοδοι μηχανικής μάθησης αντιμετωπίζουν συχνά προβλήματα από συνεχιζόμενους εξελισσόμενους κινδύνους.

Η μηχανική μάθηση εφαρμόζεται με επιτυχία σχεδόν σε οποιοδήποτε σενάριο που υπάρχουν διαθέσιμα δεδομένα από τα οποία μπορεί να εξαχθεί χρήσιμη πληροφορία. Εν τούτοις, πρέπει να εφαρμοστούν διαφορετικές τεχνικές, αναλόγως του προβλήματος. Υπάρχουν τρεις βασικές κατηγορίες που καθορίζονται από την φύση του προβλήματος και θα αναλυθούν παρακάτω:

1. Μη επιτηρούμενη μάθηση
2. Ενισχυτική μάθηση
3. Επιτηρούμενη μάθηση

Παρακάτω θα γίνει μία σύντομη αναφορά στις μεθόδους μη επιτηρούμενης και ενισχυτικής μάθησης και στη συνέχεια μία πιο εκτενής αναφορά σε μεθόδους επιτηρούμενης μάθησης, καθώς η παρούσα εργασία επικεντρώνεται σε αυτές.

### 3.2.1 Μη Επιτηρούμενη Μάθηση (Unsupervised Learning)

Ο στόχος της μάθησης χωρίς επίβλεψη είναι η εκμάθηση μιας συνάρτησης που βρίσκει μοτίβα ή τάσεις στο σύνολο δεδομένων. Για παράδειγμα, η συνάρτηση εκμάθησης θα μπορούσε να ομαδοποιεί δείγματα στο σύνολο των δεδομένων με βάση την ομοιότητά τους σύμφωνα με ορισμένα χαρακτηριστικά.

Παρακάτω, περιγράφονται οι πιο κοινές μέθοδοι μάθησης χωρίς επίβλεψη, οι οποίες συχνά χρησιμοποιούνται ως βήματα προ-επεξεργασίας:

- [k- Means Clustering](#)

Συνήθως χρησιμοποιείται ως τεχνική προεπεξεργασίας για τη μείωση του αριθμού των δεδομένων συνοψίζοντάς τα σύμφωνα με τις μέσες αναμενόμενες τιμές [6], [35]. Με άλλα



λόγια, παίρνει έναν αριθμό δειγμάτων ( $n$ ) και στοχεύει να τα χωρίσει σε ορισμένα σύνολα ( $k$ , όπου  $k < n$ ) έτσι ώστε η διακύμανση σε κάθε ομάδα να ελαχιστοποιείται. Ο πιο συνηθισμένος αλγόριθμος είναι ο «επαναληπτικός ή αφελής»  $k$ -means σύμφωνα με [36].

- [Principal Component Analysis \(PCA\)](#)

Η PCA είναι μια κοινή τεχνική προεπεξεργασίας για τη μείωση του αριθμού των χαρακτηριστικών διατηρώντας τη διακύμανσή τους [6], [36].

- [Generative Adversarial Networks \(GANs\)](#)

Τα Generative Adversarial Networks (GANs) αποτελούν μια αρχιτεκτονική στην οποία δύο δίκτυα («γεννήτρια και διαχωριστής») ανταγωνίζονται. Ο στόχος της γεννήτριας είναι να παράγει δείγματα ικανά να ξεγελάσουν τον «διαχωριστή», ενώ ο ρόλος του «διαχωριστή» είναι να ανιχνεύει ψευδή παραδείγματα από τη γεννήτρια. Με άλλα λόγια, δεδομένου ενός συνόλου εκπαίδευσης, αυτή η αρχιτεκτονική είναι σε θέση να δημιουργήσει νέα δεδομένα που μοιάζουν στατιστικά με τα αρχικά στοιχεία [6],[37].

### 3.2.2 Ενισχυτική Μάθηση (Reinforcement Learning)

Αρχικό σύνολο δεδομένων δεν υπάρχει καν εκ των προτέρων, αλλά νέα δείγματα  $x_i$  φτάνουν ή δημιουργούνται εν κινήσει, όπως η τρέχουσα κατάσταση μιας σκακίερας σε έναν αγώνα μετά την εκτέλεση μιας κίνησης. Μια συνάρτηση  $g$  που παράγει μια έξοδο  $z_i$  αντλώντας δεδομένα από ένα δείγμα και τροποποιεί την τρέχουσα κατάσταση για να παράγει ένα άλλο δείγμα  $x_{i+1}$ . Σε οποιαδήποτε κατάσταση ή σε συγκεκριμένους χρόνους, μπορούμε να μετρήσουμε πόσο καλά συμπεριφέρεται αυτό το  $g$  σύμφωνα με προκαθορισμένα κριτήρια. Ο στόχος είναι να μάθουμε ένα  $g$  που θα μεγιστοποιεί (ή θα ελαχιστοποιεί) τέτοια κριτήρια.

Η Ενισχυτική Μάθηση (RL) μπορεί τυπικά να υποδιαιρεθεί σε δύο κύριες κατηγορίες [38]. Βάσει μοντέλου και χωρίς μοντέλο. Το πρώτο χτίζει ένα εσωτερικό μοντέλο των πιθανών καταστάσεων, μεταβάσεων και αποτελεσμάτων στο περιβάλλον. Η Ενισχυτική Μάθηση χωρίς μοντέλο δεν χρησιμοποιεί κανένα μοντέλο, αλλά μάλλον μαθαίνει ενέργειες/μεταβάσεις απευθείας από την εμπειρία σε βάρος της στατιστικής αποτελεσματικότητας.

Ο πιο συνηθισμένος αλγόριθμος ενισχυτικής μάθησης για τα χρηματοοικονομικά είναι ο Q-learning [39] και η επέκτασή του σε Deep Q-learning [37]. Το Q-learning μαθαίνει μια λεγόμενη συνάρτηση ποιότητας ή συνάρτηση αξίας-δράσης, η οποία περιγράφει πόσο καλό είναι να αναλαμβάνεις μια συγκεκριμένη ενέργεια σε μια καθορισμένη κατάσταση. Για να γίνει αυτό, διατηρείται ένας πίνακας ζευγών κατάστασης-δράσεων. Αυτός ο πίνακας



καταχωρεί μια βαθμωτή «ανταμοιβή» που καθορίζει την ποιότητα της ενέργειας σε μια δεδομένη κατάσταση. Κατά τη διάρκεια της εκπαίδευσης, οι ενέργειες εκτελούνται είτε τυχαία είτε με βάση την καλύτερη στον πίνακα. Αναλύοντας την ανταμοιβή μετά από κάθε ενέργεια, ο πίνακας «κατάστασης- δράσης» μπορεί να ενημερωθεί με βάση και τις παλιές και τις νέες τιμές ανταμοιβής- επιβράβευσης. Το Deep Q-learning διατηρεί την ίδια διαδικασία, αλλά χρησιμοποιεί «βαθιά» νευρωνικά δίκτυα για να αναπαραστήσει τον πίνακα ενεργειών κατάστασης. Συνήθως εφαρμόζεται σε προβλήματα στα οποία ο χώρος επιλογής είναι τόσο μεγάλος που ο καθορισμός ενός πίνακα ενεργειών κατάστασης θα ήταν πολύ περίπλοκος και υπολογιστικά ακριβός.

Μια άλλη επιτυχημένη τάση της Ενισχυτικής Μάθησης είναι το Recurrent Reinforcement Learning (RRL) [41]. Το RRL συνδυάζει την εποπτευόμενη μάθηση με την ενισχυτική μάθηση συνήθως χρησιμοποιώντας ένα RNN για την εκμάθηση της αναπαράστασης κρυφών καταστάσεων για τον αλγόριθμο RL, ο οποίος είναι συνήθως ένα βαθύ δίκτυο Q-learning για την απόκτηση της μεθόδου που μεγιστοποιεί την ανταμοιβή.

### 3.2.3 Επιτηρούμενη Μάθηση (Supervised Learning)

Οι παρακάτω είναι οι πιο γνωστές κλασικές μέθοδοι επιτηρούμενης μάθησης που εφαρμόζονται στο πεδίο των χρηματοοικονομικών:

- [Least Squares](#)

Μία μέθοδος που τυπικά εφαρμόζεται για την εύρεση μίας γραμμικής σχέσης (παλινδρόμηση), βρίσκοντας την καλύτερη προσαρμογή στο πλαίσιο των ελαχίστων τετραγώνων, δηλαδή ελαχιστοποιώντας το άθροισμα των τετραγωνικών καταλοίπων [9].

- [LASSO και ridge regression](#)

Μια μορφή γραμμικής παλινδρόμησης που χαρακτηρίζεται για χρήση μεθόδων συρρίκνωσης [13]. Αυτό σημαίνει ότι το LASSO εκτελεί κανονικοποίηση τύπου L1 για να ελαχιστοποιήσει την απόλυτη τιμή του μεγέθους των συντελεστών παλινδρόμησης. Ως αποτέλεσμα, συνήθως παράγεται ένα «αραιό» σύνολο συντελεστών, συμβάλλοντας στη μείωση της υπερπροσαρμογής και της πολυπλοκότητας του μοντέλου. Η παλινδρόμηση ridge λειτουργεί με παρόμοιο τρόπο, εκτελώντας κανονικοποίηση τύπου L2, η οποία δεν παράγει αραιά μοντέλα.

Η ridge regression προσθέτει το άθροισμα των τετραγώνων των συντελεστών παλινδρόμησης ως όρο ποινής στη συνάρτηση απώλειας, ως εξής:

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

Αντίθετα η παλινδρόμηση Lasso (Least Absolute Shrinkage and Selection Operator) χρησιμοποιεί το άθροισμα των απόλυτων τιμών των συντελεστών παλινδρόμησης ως όρο ποινής στη συνάρτηση απώλειας:

$$\sum_{i=1}^n (Y_i - \sum_{j=1}^p X_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

- [Decision Trees](#)

Ένα δέντρο αποφάσεων είναι ένας από τους απλούστερους αλλά ισχυρούς αλγόριθμους μηχανικής μάθησης για εργασίες ταξινόμησης και παλινδρόμησης [11]. Είναι μια αρχιτεκτονική μηχανικής μάθησης που χρησιμοποιεί μια δομή παρόμοια με το διάγραμμα ροής για να καταλήξει σε ένα συμπέρασμα μέσω μιας σειράς ελέγχων για τις μεταβλητές εισόδου. Κάθε κόμβος του δέντρου είναι ένας έλεγχος σε μια μεταβλητή εισόδου και ανάλογα με το αποτέλεσμα, η ροή συνεχίζεται σε έναν κλάδο του δέντρου ή στον άλλο μέχρι η ροή να φτάσει στα φύλλα όπου δίνονται οι τελικές έξοδοι. Ο αλγόριθμος επιχειρεί να προσδιορίσει τη διαδικασία απόφασης για την εργασία ταξινόμησης από τα δεδομένα: κατασκευάζει πρώτα μια συνάρτηση κόστους. Στη συνέχεια, ο αλγόριθμος ξεκινά εξετάζοντας πρώτα όλα τα χαρακτηριστικά και δοκιμάζει διαφορετικά σημεία διαχωρισμού για κάθε χαρακτηριστικό για να εντοπίσει τον πιο αποδοτικό διαχωρισμό. Για κάθε έναν από τους κλάδους του διαχωρισμού σε αυτό το επίπεδο, επαναλαμβάνεται η ίδια διαδικασία διαχωρισμού για τα υπόλοιπα χαρακτηριστικά. Ο αλγόριθμος σταματά όταν επιτευχθεί ο προκαθορισμένος αριθμός επιπέδων (που ονομάζεται βάθος). Το βάθος είναι μια υπερπαράμετρος και μπορεί να ρυθμιστεί για να ελέγχει τη διαδικασία εκμάθησης.

- [Random Forest](#)

Ο αλγόριθμος Random Forest είναι μια μέθοδος εκμάθησης συνόλου που βασίζεται στα δέντρα αποφάσεων [11],[13],[14],[19]. Σε ένα δέντρο αποφάσεων, το δείγμα χωρίζεται αναδρομικά σε πολλές ομοιογενείς και μη επικαλυπτόμενες περιοχές με βάση τα πιο σχετικά χαρακτηριστικά ή προγνωστικούς παράγοντες. Τα δέντρα αποφάσεων είναι εξαιρετικά ερμηνεύσιμα και επιλέγουν αυτόματα τις μεταβλητές πρόβλεψης χωρίζοντας το δείγμα σε κάθε κόμβο. Ωστόσο, η απόδοση πρόβλεψής τους μπορεί να είναι κακή λόγω της μεγάλης διακύμανσης των προβλέψεων.

Ο αλγόριθμος Random Forest αντί να κατασκευάσει ένα δέντρο αποφάσεων, κατασκευάζει πολλαπλά δέντρα απόφασης μόνο με βάση τυχαία επιλεγμένα υποσύνολα χαρακτηριστικών (εναλλακτικά, η τυχαιοποίηση μπορεί επίσης να εφαρμοστεί σε δεδομένα). Στη συνέχεια, χρησιμοποιείται ένας συνδυασμός των επιμέρους δέντρων για να εξαχθεί ένα τελικό μοντέλο, εξ ου και το όνομα «Random Forest (Τυχαίο Δάσος)». Το βάθος κάθε δέντρου απόφασης και ο αριθμός των δέντρων απόφασης είναι υπερπαραμέτροι και μπορούν να ρυθμιστούν για τη βελτίωση της μάθησης. Ο αλγόριθμος Random Forest αποφεύγει το πρόβλημα υπερπροσαρμογής ενός μεμονωμένου δέντρου αποφάσεων μέσω του συνδυασμού πολλών δέντρων, καθιστώντας το τελικό αποτέλεσμα πιο ακριβές αλλά λιγότερο ερμηνεύσιμο λόγω της παρουσίας πολλών μικρών δέντρων. Πολλές εργασίες και έρευνες έχουν δείξει ότι η μέθοδος Random Forest παρέχει αποτελέσματα υψηλής ακρίβειας ειδικά αν το πρόβλημα έχει πολλές διαστάσεις και η συσχέτιση μεταξύ των μεταβλητών είναι μη γραμμική με πολλές αλληλεπιδράσεις μεταξύ τους [17], [18].

Η μέθοδος Random Forest βασίζεται σε μία τεχνική που ονομάζεται «Bagging» ή bootstrap aggregation. Η τεχνική αυτή στην ουσία, είναι μια τεχνική για την μείωση της διακύμανσης μιας εκτιμώμενης συνάρτησης πρόβλεψης. Η τεχνική «bagging» φαίνεται να λειτουργεί ιδιαίτερα καλά για διαδικασίες υψηλής διακύμανσης και χαμηλής προκατάληψης (bias), όπως τα δέντρα αποφάσεων.

Η μέθοδος Random Forest μειώνει τη διακύμανση των προβλέψεων των δέντρων απόφασης. Το αποτέλεσμα του αλγορίθμου Random Forest βασίζεται στη μέση πρόβλεψη για όλα τα δέντρα που συνδυάζονται. Η μείωση της διακύμανσης της πρόβλεψης σχετίζεται με τον βαθμό ανεξαρτησίας (συσχετίσεις) μεταξύ των μεμονωμένων δέντρων, και λόγω αυτού τα δέντρα θα πρέπει να είναι όσο το δυνατόν λιγότερο συσχετισμένα. Η εν λόγω μέθοδος μπορεί να χρησιμοποιηθεί τόσο σε προβλήματα με δέντρα ταξινόμησης όσο και με δέντρα παλινδρόμησης.

Η μέθοδος αυτή καθώς και μία επέκτασή της η Ordinal Random Forest [16] χρησιμοποιήθηκαν στο πειραματικό σκέλος της εργασίας, που θα αναλυθεί στο επόμενο κεφάλαιο.

- [Gradient Boosting](#)

Ο αλγόριθμος Gradient Boosting είναι μια τεχνική μηχανικής μάθησης για προβλήματα παλινδρόμησης και ταξινόμησης, η οποία παράγει ένα μοντέλο πρόβλεψης το οποίο συνδυάζει πολλά επιμέρους μοντέλα πρόβλεψης, συνήθως δέντρα αποφάσεων [19], [21], [22], [23]. Ο αλγόριθμος κατασκευάζει το τελικό μοντέλο σταδιακά, όπως κάνουν άλλες μέθοδοι ενίσχυσης. Η ιδέα της ενίσχυσης κλίσης (gradient boosting) προήλθε από την παρατήρηση του

Leo Breiman ότι η ενίσχυση μπορεί να ερμηνευτεί ως ένας αλγόριθμος βελτιστοποίησης σε μια κατάλληλη συνάρτηση κόστους [22].

Αντί να συγκεντρώνει δέντρα απόφασης με ανεξάρτητο τρόπο, όπως στην περίπτωση των Random Forest, ο αλγόριθμος Gradient Boosting βασίζεται σε ένα σταδιακό συνδυασμό των δέντρων, ξεκινώντας από δέντρα που παρέχουν ασθενή αποτελέσματα (αποτελέσματα που είναι ελαφρώς καλύτερα από μία τυχαία εικασία) και προχωρώντας σταδιακά με δέντρα υψηλότερης ακρίβειας. Με αυτόν τον τρόπο, η ενίσχυση μπορεί να επιτύχει βελτιωμένες προβλέψεις μειώνοντας όχι μόνο την προκατάληψη πρόβλεψης αλλά και τη διακύμανση [23].

Το boosting μαθαίνει πως να συγκεντρώνει τα δέντρα απόφασης σταδιακά προκειμένου να δώσει μεγαλύτερη έμφαση σε παρατηρήσεις οι οποίες δεν έχουν αξιολογηθεί με επιτυχία από τα προηγούμενα δέντρα. Η ενίσχυση κλίσης στοχεύει στη βελτίωση της ακρίβειας των προβλέψεων χρησιμοποιώντας μια συνάρτηση απώλειας η οποία ελαχιστοποιείται (χρησιμοποιώντας την κατάβαση κλίσης) προσθέτοντας το σφάλμα πρόβλεψης των δέντρων διαδοχικά. Ως εκ τούτου, η ενίσχυση κλίσης είναι σε θέση να αναγνωρίσει μεγάλα σφάλματα από προηγούμενες επαναλήψεις ελαχιστοποιώντας τη συνάρτηση απώλειας, η οποία είναι συνήθως το μέσο τετραγωνικό σφάλμα των προβλέψεων σε ένα πλαίσιο παλινδρόμησης. Συγκεκριμένα, η συνάρτηση πρόβλεψης ενημερώνεται στην επανάληψη  $b$  ως:

$$F_{b+1}(z_t) = F_b(z_t) - \delta_b h_b(z_t)$$

όπου το  $F$  υποδηλώνει τη συνάρτηση πρόβλεψης, το  $h$  είναι ένα δέντρο που υπολογίζεται από υπολείμματα κλίσης, και το  $\delta$  είναι ο ρυθμός εκμάθησης (υπερ-παράμετρος).

Σε αντίθεση με τον αλγόριθμο Random Forest, η Gradient Boosting τείνει να παρουσιάζει προβλήματα από την υψηλή προσαρμογή στα δεδομένα εκμάθησης. Για να αποφευχθεί το πρόβλημα αυτό, χρησιμοποιούνται διάφορες τεχνικές, όπως: περιορισμοί στα χαρακτηριστικά των αναπτυσσόμενων δέντρων (αριθμός δέντρων, βάθος δέντρου, αριθμός κόμβων, κ.λπ.), συρρίκνωση του ρυθμού εκμάθησης, τυχαία υποδειγματοληψία των δεδομένων (χωρίς αντικατάσταση), αποκλεισμός τιμών σε τερματικούς κόμβους (όπως στον αλγόριθμο XGboost).

- **XGBoost (eXtreme Gradient Boosting)**

Το XGBoost (eXtreme Gradient Boosting) [21],[24] είναι μια βιβλιοθήκη λογισμικού ανοιχτού κώδικα για την ανάπτυξη μοντέλων μηχανικής μάθησης, μέσω μιας εξελιγμένης υλοποίησης της ιδέας του gradient boosting. Έχει κερδίσει μεγάλη δημοτικότητα και προσοχή ως ο αλγόριθμος επιλογής για πολλές νικήτριες ομάδες διαγωνισμών μηχανικής μάθησης.

Το XGBoost ξεκίνησε αρχικά ως ερευνητικό έργο από τον Tianqi Chen [25] ως μέρος της ομάδας Distributed (Deep) Machine Learning Community (DMLC). Έγινε πολύ γνωστό στους κύκλους των διαγωνισμών μηχανικής μάθησης μετά τη χρήση του στη νικήτρια λύση του Higgs Machine Learning Challenge. Αμέσως μετά, δημιουργήθηκαν τα πακέτα Python και R και το XGBoost έχει πλέον υλοποιήσεις πακέτων για Java, Scala, Julia, Perl και άλλες γλώσσες. Αυτό έφερε τη βιβλιοθήκη σε περισσότερους προγραμματιστές και συνέβαλε στη δημοτικότητα της στην κοινότητα Kaggle, όπου έχει χρησιμοποιηθεί για μεγάλο αριθμό διαγωνισμών. Σύντομα ενσωματώθηκε σε μια σειρά από άλλα πακέτα που διευκολύνουν τη χρήση στις αντίστοιχες κοινότητές τους. Έχει πλέον ενσωματωθεί στο scikit-learn για χρήστες Python και στο πακέτο Caret για χρήστες R. Μπορεί επίσης να ενσωματωθεί σε πλαίσια δεδομένων ροής όπως το Apache Spark, το Apache Hadoop και το Apache Flink .

Ενώ το μοντέλο XGBoost συχνά επιτυγχάνει υψηλότερη ακρίβεια από ένα μεμονωμένο δέντρο αποφάσεων, θυσιάζει την εγγενή ερμηνευσιμότητα των δέντρων αποφάσεων. Για παράδειγμα, ενώ είναι εύκολη η κατανόηση του μονοπατιού που ακολουθεί ένα δέντρο αποφάσεων για την εξαγωγή του τελικού αποτελέσματος, αυτό είναι αδύνατο σε ένα μοντέλο που συνδυάζει πολλά δέντρα.

- [Λογιστική παλινδρόμηση](#)

Στη στατιστική, το λογιστικό μοντέλο μοντελοποιεί την πιθανότητα να λάβει χώρα ένα γεγονός, έχοντας τις πιθανότητες καταγραφής για το γεγονός, ως ένας γραμμικός συνδυασμός ανεξάρτητων μεταβλητών. Στην ανάλυση παλινδρόμησης, η λογιστική παλινδρόμηση [26], [27] (ή παλινδρόμηση logit) είναι η εκτίμηση των παραμέτρων ενός λογιστικού μοντέλου (δηλαδή των συντελεστών στον γραμμικό συνδυασμό). Ένα μοντέλο λογιστικής παλινδρόμησης εκφράζει μία γραμμική σχέση μεταξύ ανεξάρτητων μεταβλητών και μιας εξαρτημένης μεταβλητής. Σημαντικό ρόλο παίζει η επιλογή των μεταβλητών. Η επιλογή μεταβλητών στοχεύει στη μείωση του αριθμού των μεταβλητών σε ένα μοντέλο. Θα κάνει το μοντέλο πιο συνοπτικό και επομένως πιο ερμηνεύσιμο, πιο γρήγορο για αξιολόγηση, και πιο αξιόπιστο μειώνοντας τη συγγραμμικότητα.

Σε πολλές στατιστικές μελέτες και αναλύσεις εταιρειών- πελατών, είναι πολύ σημαντικό να γνωρίζουμε ότι η στατιστική σημασία είναι μόνο ένα κριτήριο αξιολόγησης για την επιλογή μεταβλητών. Όπως αναφέρθηκε προηγουμένως, η ερμηνευτικότητα είναι επίσης ένα σημαντικό κριτήριο. Τόσο στη γραμμική όσο και στη λογιστική παλινδρόμηση, αυτό μπορεί εύκολα να αξιολογηθεί εξετάζοντας τα πρόσημα των συντελεστών παλινδρόμησης.

- Μονότονη λογιστική παλινδρόμηση

Η μονότονη λογιστική παλινδρόμηση (ordinal logistic regression, OLR [28], [30]) είναι μια μέθοδος στατιστικής ανάλυσης που μπορεί να χρησιμοποιηθεί για τη μοντελοποίηση της σχέσης μεταξύ μιας μεταβλητής τακτικής απόκρισης (ordinal response) και μιας ή περισσότερων επεξηγηματικών μεταβλητών. Τακτική μεταβλητή είναι μια κατηγορική μεταβλητή για την οποία υπάρχει σαφής σειρά και διάκριση των επιπέδων της κάθε κατηγορίας. Οι επεξηγηματικές μεταβλητές μπορεί να είναι είτε συνεχείς είτε κατηγορικές. Η εκτίμηση των μοντέλων OLR με στατιστικό λογισμικό δεν είναι δύσκολη, αλλά η ερμηνεία της παραγωγής του μοντέλου μπορεί να είναι αρκετά περίπλοκη.

Η OLR είναι μια επέκταση της λογιστικής παλινδρόμησης όπου η σχέση μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών είναι γραμμική. Μια σημαντική υπόθεση της OLR είναι η υπόθεση των αναλογικών πιθανοτήτων: η επίδραση μιας ανεξάρτητης μεταβλητής είναι σταθερή για κάθε αύξηση του επιπέδου της απόκρισης. Ως εκ τούτου, η έξοδος μιας OLR θα περιέχει μια τομή για κάθε επίπεδο της απόκρισης εκτός από ένα, και μια μόνο κλίση για κάθε επεξηγηματική μεταβλητή.

- Νευρωνικά Δίκτυα (Neural Networks- NNs)

Τα νευρωνικά δίκτυα (NN) που συνήθως αποτελούνται από μεμονωμένα perceptrons (συστήματα αντίληψης) τα οποία είναι διατεταγμένα σε πολλαπλά στρώματα [6], [31]. Ένα perceptron [31] είναι μια συνάρτηση  $f$  που χαρτογραφεί μια είσοδο  $x$  για να δημιουργήσει μια έξοδο  $z$  με τον ακόλουθο τρόπο:

$$z = f(x) = \sigma(wx + b),$$

όπου  $w$  είναι ένα διάνυσμα βαρών,  $b$  είναι μια σταθερά και  $\sigma$  είναι μια συνάρτηση ενεργοποίησης. Στην πιο απλή μορφή της, η συνάρτηση ενεργοποίησης είναι απλώς ένα κατώφλι και το perceptron είναι απλώς ένας δυαδικός ταξινομητής. Η σταθερά απλώς μετατοπίζει το όριο απόφασης μακριά από την προέλευση. Τα μονοστρωματικά perceptron μπορούν να συνδυαστούν για να σχηματίσουν ένα perceptron πολλαπλών στρωμάτων (multi-layer perceptron, MLP). Αυτή η αρχιτεκτονική αποτελείται συνήθως από τρία επίπεδα: το επίπεδο εισόδου, ένα κρυφό επίπεδο και ένα επίπεδο εξόδου. Το κρυφό επίπεδο και το επίπεδο εξόδου μπορούν να αποτελούνται από έναν αυθαίρετο αριθμό κόμβων (που ονομάζονται επίσης νευρώνες). Καθένας από αυτούς τους κόμβους είναι ένα perceptron μονής στρώσης που χρησιμοποιεί μια μη γραμμική συνάρτηση ενεργοποίησης. Τα Deep Neural Network (Deep-NN) (ονομάζονται επίσης πλήρως συνδεδεμένα δίκτυα) συχνά αναφέρονται σαν MLP με περισσότερα κρυφά επίπεδα. Σε αυτή τη γενική περίπτωση, η έξοδος ενός συγκεκριμένου νευρώνα  $i$  ενός στρώματος  $l$  μπορεί να οριστεί ως εξής:

$$z_i^l = \sigma \sum_k w_k^l x_{i-k}^{l-1}$$

Τα “Vanilla NNs” είναι ικανά να μάθουν οποιαδήποτε μη γραμμική συνάρτηση (χαρακτηρίζονται ως «καθολικοί προσεγγιστές») δεδομένης της αρκετής πολυπλοκότητας του δικτύου, αλλά αντιμετωπίζουν μια σειρά από προκλήσεις:

- (1) λόγω της πλήρως συνδεδεμένης φύσης τους, απαιτούν μεγάλο αριθμό παραμέτρων,
- (2) είναι συνήθως πιο δύσκολη η εκπαίδευσή τους,
- (3) χάνουν τις χωρικές πληροφορίες της εισόδου και
- (4) δεν υπάρχει ενσωματωμένος μηχανισμός για την αξιοποίηση των νέων δεδομένων.

- [Convolutional Neural Networks \(CNNs\)](#)

Τα Convolutional Neural Networks (CNNs, [6], [32]) χρησιμοποιούν φίλτρα πυρήνα με δυνατότητα εκμάθησης για να εξάγουν τα σχετικά χαρακτηριστικά από τις εισόδους εφαρμόζοντας τη λειτουργία της συνέλιξης με αυτά. Είναι ιδιαίτερα χρήσιμα με δομημένα δεδομένα και σε εκείνες τις περιπτώσεις όπου οι χωρικές πληροφορίες είναι σημαντικές. Συνήθως, εφαρμόζονται επί του παρόντος για την επεξεργασία εικόνων 2D (αν και μια συνέλιξη μπορεί να εφαρμοστεί σε οποιαδήποτε διάσταση). Ένα CNN μοιράζεται αυτόν τον πυρήνα σε ολόκληρο το επίπεδο και, με αυτόν τον τρόπο, δεν κάνει συνδέει πλήρως κάθε νευρώνα από το προηγούμενο στρώμα με τους επόμενους. Επιπλέον, κάθε επίπεδο μπορεί να έχει περισσότερους από έναν πυρήνες. Το καθένα περιπλέκεται ξεχωριστά για να παράγει ξεχωριστή έξοδο. Αυτές οι έξοδοι αναφέρονται συχνά ως χάρτες χαρακτηριστικών και στοιβάζονται για να σχηματίσουν μια έξοδο πολλαπλών καναλιών.

Όσον αφορά τα πλήρως συνδεδεμένα NN, παρουσιάζουν ορισμένα πλεονεκτήματα:

1. συνδυάζοντας την είσοδο με φίλτρα προκαθορισμένου μεγέθους αντί να είναι πλήρως συνδεδεμένα, καταγράφουν χωρικά χαρακτηριστικά και
2. επειδή δεν είναι πλήρως συνδεδεμένα, αλλά αντίθετα μοιράζονται τα βάρη του πυρήνα σε ολόκληρη την είσοδο, απαιτούν πολύ λιγότερες παραμέτρους και επομένως είναι ευκολότερο να εκπαιδεύονται και λιγότερο επιρρεπή σε τυχόν υπερπροσαρμογή (overfitting).

- [Recurrent Neural Networks \(RNNs\)](#)

Τα Recurrent Neural Networks (RNNs) είναι ειδικά σχεδιασμένα για να αντιμετωπίζουν δεδομένα ακολουθίας και να μαθαίνουν από χρονικές πληροφορίες [6], [33], [34]. Αν και



εσωτερικά μπορούν να διαμορφωθούν είτε ως παραδοσιακά NN είτε ως CNN, συνήθως προσθέτουν επαναλαμβανόμενες συνδέσεις στα στρώματά τους, κάτι που βοηθά να ληφθεί υπόψη η κατάσταση από προηγούμενα στοιχεία ακολουθίας ή χρονικές στιγμές. Επομένως, μπορούν να καταγράψουν διαδοχικές πληροφορίες και να μοιράζονται παραμέτρους σε διαφορετικά χρονικά βήματα (με παρόμοιο τρόπο όπως κάνουν τα CNN χωρικά). Τυπικά, η πιο γενική τοπολογία είναι ένα πλήρως επαναλαμβανόμενο RNN όπου οι έξοδοι όλων των νευρώνων συνδέονται με τις εισόδους όλων για αυτούς. Κάθε ένα πολλαπλασιάζει τις τρέχουσες εισόδους και τις προηγούμενες εξόδους μέσω μιας συνάρτησης ενεργοποίησης. Άλλες σχετικές τοπολογίες είναι το Gated Recurrent Unit Network (GRU) και η ευρέως διαδεδομένη Long- Short Term Memory (LSTM).

Τα GRUs [33] διαθέτουν δύο μηχανισμούς πύλης: πύλη ενημέρωση και πύλη επαναφορά. Η πύλη ενημέρωσης είναι υπεύθυνη για τον προσδιορισμό του όγκου των προηγούμενων πληροφοριών που θα ρέουν στο επόμενο βήμα. Η πύλη επαναφοράς αποφασίζει ποιες πληροφορίες από το προηγούμενο χρονικό βήμα θα παραληφθούν για την τρέχουσα κατάσταση.

Οι LSTMs [34] διαθέτουν τρεις μηχανισμούς πύλης: είσοδο, έξοδο και λήθη. Αυτό το σύστημα τριπλής πύλης επιτρέπει στην αρχιτεκτονική να μοντελοποιεί σωστά τις μακροπρόθεσμες και βραχυπρόθεσμες εξαρτήσεις.

Στην πραγματικότητα, όλα τα Vanilla RNN, GRU ή LSTM είναι σε θέση να μοντελοποιούν αυθαίρετες χρονικές εξαρτήσεις. Το πρόβλημα είναι, ωστόσο, υπολογιστικό και αριθμητικό: λόγω της φύσης της εκπαιδευτικής διαδικασίας, οι απαιτούμενες κλίσεις για μάθηση μπορούν εύκολα να στραφούν στο άπειρο ή να εξαφανιστούν (να πάνε στο μηδέν) αποτρέποντας οποιαδήποτε μάθηση. Τα GRU είναι ένα βήμα μπροστά σε σύγκριση με τα Vanilla RNN και οι πρόσθετες πύλες από τα LSTM βοηθούν ακόμη περισσότερο στον έλεγχο της ροής πληροφοριών για την αποφυγή αυτών των προβλημάτων.



## 4. ΠΕΙΡΑΜΑΤΙΚΟ- ΕΡΕΥΝΗΤΙΚΟ ΣΚΕΛΟΣ

Σκοπός του ερευνητικού τμήματος της εργασίας είναι να διαπιστώσουμε ποια από τις μεθόδους που εφαρμόσαμε για την αξιολόγηση αμοιβαίων κεφαλαίων, παρέχει καλύτερα αποτελέσματα στην εκτίμηση της αξιολόγησης των αμοιβαίων κεφαλαίων. Ως βάση- πρότυπη αξιολόγηση χρησιμοποιήθηκε η αξιολόγηση της εταιρείας Morningstar η οποία ταξινομεί τα αμοιβαία κεφάλαια σε πέντε κατηγορίες λαμβάνοντας υπόψιν της μια σειρά παραγόντων και μεταβλητών.

### 4.1 Εκκαθάριση- Προεπεξεργασία Δεδομένων

Η πηγή των δεδομένων που χρησιμοποιήθηκαν ήταν από το Kaggle.com. Αρχικά η λίστα περιείχε 23783 διαφορετικά αμοιβαία κεφάλαια τα οποία δραστηριοποιούνται στις ΗΠΑ. Για κάθε αμοιβαίο κεφάλαιο υπήρχαν αρχικά 297 μεταβλητές, ποιοτικές και ποσοτικές. Λόγω των ελλিপών στοιχείων, τελικά χρησιμοποιήθηκαν δεδομένα για 9410 αμοιβαία κεφάλαια.

Στο επόμενο στάδιο της προεπεξεργασίας των δεδομένων έγινε η επιλογή των κατάλληλων μεταβλητών για την πραγματοποίηση της ανάλυσης. Αρχικά διαπιστώθηκε ότι αρκετές μεταβλητές είχαν ισχυρή και άμεση συσχέτιση μεταξύ τους. Έτσι έγινε συγχώνευση κάποιων μεταβλητών με αποτέλεσμα να διαμορφωθεί σε μία λίστα με 45 μεταβλητές όπως φαίνεται στον παρακάτω πίνακα.

Πίνακας 4.1: Μεταβλητές της ανάλυσης

ΜΕΤΑΒΛΗΤΗ	ΠΕΡΙΓΡΑΦΗ
fund_category	Κατηγορία κεφαλαίου
overall_rating	Συνολική ταξινόμηση Morningstar
risk_rating	Ταξινόμηση βάσει κινδύνου
Turnover	Ανακύκλωση επενδύσεων κεφαλαίου
investment_type	Κατηγορία επενδύσεων
size_type	Κατηγορία μεγέθους
ER	Δείκτης καθαρών εξόδων
cat_ER	Δείκτης καθαρών εξόδων της κατηγορίας του κεφαλαίου
HHI	Δείκτης Herfindahl- Hirschman
HHI_sector	Δείκτης HHI ανά κλάδο επενδύσεων
BR	Δείκτης τιμής / λογιστική αξία
cat_BR	Δείκτης τιμής / λογιστική αξία της κατηγορίας του κεφαλαίου
CR	Δείκτης τιμής / ταμειακών ροών
cat_CR	Δείκτης τιμής / ταμειακών ροών της κατηγορίας του κεφαλαίου
PE	Δείκτης τιμής / κέρδη

cat_PE	Δείκτης τιμής / κέρδη της κατηγορίας του κεφαλαίου
PS	Δείκτης τιμής / πωλήσεις
cat_PS	Δείκτης τιμής / πωλήσεις της κατηγορίας του κεφαλαίου
CAP	Δείκτης διαμέσου / κεφαλαιοποίηση
cat_CAP	Δείκτης διαμέσου / κεφαλαιοποίηση κατηγορίας κεφαλαίου
EarningsGrowth	Δείκτης ανάπτυξης κερδών τριετίας
cat_EarningsGrowth	Δείκτης ανάπτυξης κερδών τριετίας κατηγορίας κεφαλαίου
top10assets	Δείκτης δέκα σημαντικότερων επενδύσεων/ σύνολο επενδύσεων
return_rating	Αξιολόγηση Morningstar με βάση τα κέρδη
ret1year	Απόδοση ενός έτους
ret1year_cat	Απόδοση ενός έτους κατηγορίας κεφαλαίου
cat_ret1year	Διαφορά απόδοσης ενός έτους κεφαλαίου σε σχέση με τα κεφάλαια της κατηγορίας
ret3years	Απόδοση τριών ετών
ret3year_cat	Απόδοση τριών ετών κατηγορίας κεφαλαίου
cat_ret3years	Διαφορά απόδοσης τριών ετών κεφαλαίου σε σχέση με τα κεφάλαια της κατηγορίας
ret5years	Απόδοση πέντε ετών
ret5year_cat	Απόδοση πέντε ετών κατηγορίας κεφαλαίου
cat_ret5years	Διαφορά απόδοσης πέντε ετών κεφαλαίου σε σχέση με τα κεφάλαια της κατηγορίας
esg_score	Βαθμολογία ESG
peer_esg	Βαθμολογία ESG ανταγωνιστικών κεφαλαίων
esg_ratio	Δείκτης ESG/ ανταγωνιστικά κεφάλαια
env_score	Βαθμολογία περιβαλλοντικής επίδοσης
peer_environment_avg	Μέση βαθμολογία περιβαλλοντικής επίδοσης ανταγωνιστικών κεφαλαίων
env_ratio	Δείκτης περιβαλλοντικής επίδοσης/ ανταγωνιστικά κεφάλαια
social_score	Βαθμολογία κοινωνικής επίδοσης
peer_social_avg	Βαθμολογία κοινωνικής επίδοσης ανταγωνιστικών κεφαλαίων
social_ratio	Δείκτης κοινωνικής επίδοσης / ανταγωνιστικά κεφάλαια
governance_score	Βαθμολογία επίδοσης εταιρικής διακυβέρνησης
peer_governance_avg	Βαθμολογία επίδοσης εταιρικής διακυβέρνησης ανταγωνιστικών κεφαλαίων
governance_ratio	Δείκτης επίδοσης με βάση την εταιρική διακυβέρνηση / ανταγωνιστικά κεφάλαια

Στην συνέχεια πραγματοποιήθηκαν τρεις στατιστικοί έλεγχοι για την εξέταση της σημαντικότητας όλων των μεταβλητών. Εφαρμόστηκε η ANOVA (Analysis of Variance), καθώς και ο έλεγχος Kruskal-Wallis που χρησιμοποιείται στις μη κανονικές κατανομές. Οι έλεγχοι πραγματοποιήθηκαν για την διερεύνηση της σημαντικότητας των μεταβλητών σε σχέση με την συνολική ταξινόμηση των κεφαλαίων από την Morningstar (overall rating), καθώς και σε σχέση με τις αξιολογήσεις ως προς την απόδοση (return rating) και τον κίνδυνο (risk rating). Επιπλέον εξετάστηκαν οι συσχετίσεις των μεταβλητών. Τα αποτελέσματα για την κάθε κατηγορία συνοψίζονται στους παρακάτω πίνακες:

Πίνακας 4.2: Σημαντικότητα μεταβλητών σε σχέση με την συνολική αξιολόγηση των αμοιβαίων κεφαλαίων (έλεγχοι ANOVA και Kruskal- Wallis)

	K-W statistic	p-value	F statistic	p-value
turnover	164.36	0.000	42.70	0.000
ER	632.51	0.000	699.26	0.000
cat_ER	2.67	0.615	0.02	0.892
HHI	50.32	0.000	10.72	0.001
HHI_sector	65.54	0.000	9.54	0.002
BR	361.46	0.000	311.52	0.000
cat_BR	1008.12	0.000	1195.08	0.000
CR	515.84	0.000	554.10	0.000
cat_CR	1181.54	0.000	1429.73	0.000
PE	421.44	0.000	432.18	0.000
cat_PE	896.81	0.000	1042.65	0.000
PS	479.30	0.000	345.51	0.000
cat_PS	869.58	0.000	862.47	0.000
CAP	222.39	0.000	221.47	0.000
cat_CAP	475.50	0.000	461.74	0.000
EarningsGrowth	264.19	0.000	231.22	0.000
cat_EarningsGrowth	515.48	0.000	458.15	0.000
top10assets	6.57	0.161	0.95	0.330
ret1year	1143.55	0.000	1259.15	0.000
cat_ret1year	3314.88	0.000	3923.47	0.000
ret3years	1598.45	0.000	1462.41	0.000
cat_ret3years	5468.63	0.000	8810.18	0.000
ret5years	1560.50	0.000	1532.35	0.000
cat_ret5years	5943.44	0.000	9632.92	0.000
esg_score	62.40	0.000	37.76	0.000
esg_ratio	133.22	0.000	138.25	0.000
env_score	85.44	0.000	21.48	0.000
env_ratio	168.13	0.000	27.70	0.000
social_score	68.92	0.000	18.91	0.000
social_ratio	102.78	0.000	29.37	0.000
governance_score	84.88	0.000	18.07	0.000
governance_ratio	113.96	0.000	23.33	0.000

Τα αποτελέσματα και των δύο ελέγχων «συμφωνούσαν» στο κομμάτι της σημαντικότητας των μεταβλητών που σχετίζονται με την εξαρτημένη μεταβλητή «Overall Rating». Και από τις δύο μεθόδους καταλήγουμε στο συμπέρασμα ότι οι μεταβλητές «cat\_ER» και «top10assets» δεν είναι στατιστικά σημαντικές ακόμη και για επίπεδο σημαντικότητας της τάξης του 10%.

Πίνακας 4.3: Σημαντικότητα μεταβλητών για την αξιολόγηση σε σχέση με την απόδοση  
(έλεγχοι ANOVA και Kruskal- Wallis)

	K-W statistic	p-value	F statistic	p-value
turnover	118.85	0.000	29.95	0.000
ER	613.58	0.000	666.13	0.000
cat_ER	3.49	0.480	4.81	0.028
HHI	54.37	0.000	13.05	0.000
HHI_sector	67.56	0.000	3.96	0.047
BR	232.98	0.000	245.39	0.000
cat_BR	666.77	0.000	936.16	0.000
CR	382.44	0.000	442.38	0.000
cat_CR	848.36	0.000	1105.73	0.000
PE	323.47	0.000	356.71	0.000
cat_PE	714.21	0.000	887.68	0.000
PS	322.83	0.000	262.42	0.000
cat_PS	564.23	0.000	676.73	0.000
CAP	159.50	0.000	156.91	0.000
cat_CAP	240.34	0.000	270.05	0.000
EarningsGrowth	189.62	0.000	166.13	0.000
cat_EarningsGrowth	388.46	0.000	329.35	0.000
top10assets	32.20	0.000	0.02	0.890
ret1year	1214.42	0.000	1388.97	0.000
cat_ret1year	3634.29	0.000	4566.77	0.000
ret3years	1461.71	0.000	1356.96	0.000
cat_ret3years	5105.18	0.000	8060.13	0.000
ret5years	1527.48	0.000	1519.67	0.000
cat_ret5years	5998.45	0.000	10107.39	0.000
esg_score	45.16	0.000	24.66	0.000
esg_ratio	77.07	0.000	46.96	0.000
env_score	70.67	0.000	20.25	0.000
env_ratio	166.19	0.000	27.49	0.000
social_score	39.23	0.000	14.10	0.000
social_ratio	66.03	0.000	22.10	0.000
governance_score	84.01	0.000	22.39	0.000
governance_ratio	108.47	0.000	23.59	0.000

Το αποτέλεσμα του ελέγχου KW έδειξαν ότι η μεταβλητή «cat\_ER» δεν είναι στατιστικά σημαντική για τη διάκριση των κεφαλαίων ως προς την αξιολόγηση της Morningstar για την απόδοση, ενώ αντίστοιχα ο έλεγχος ANOVA έδειξε ότι η μεταβλητή «top10assets» δεν είναι στατιστικά σημαντική.

Πίνακας 4.4: Σημαντικότητα μεταβλητών για την αξιολόγηση αμοιβαίων κεφαλαίων βάσει κινδύνου (έλεγχοι ANOVA και Kruskal- Wallis)

	K-W statistic	p-value	F statistic	p-value
turnover	92.37	0.000	2.57	0.109
ER	38.56	0.000	4.97	0.026
cat_ER	41.25	0.000	25.85	0.000
HHI	192.88	0.000	2.06	0.152
HHI_sector	80.20	0.000	8.04	0.005
BR	161.19	0.000	9.25	0.002
cat_BR	224.71	0.000	14.02	0.000
CR	85.88	0.000	16.43	0.000
cat_CR	118.70	0.000	6.00	0.014
PE	60.33	0.000	12.99	0.000
cat_PE	45.94	0.000	2.66	0.103
PS	158.64	0.000	37.07	0.000
cat_PS	193.44	0.000	18.71	0.000
CAP	71.30	0.000	42.99	0.000
cat_CAP	346.66	0.000	133.08	0.000
EarningsGrowth	53.00	0.000	4.68	0.031
cat_EarningsGrowth	32.34	0.000	1.72	0.189
top10assets	67.71	0.000	0.73	0.394
ret1year	75.37	0.000	84.55	0.000
cat_ret1year	256.04	0.000	354.11	0.000
ret3years	59.40	0.000	0.19	0.660
cat_ret3years	50.37	0.000	33.81	0.000
ret5years	42.52	0.000	14.44	0.000
cat_ret5years	124.88	0.000	172.19	0.000
esg_score	86.52	0.000	61.94	0.000
esg_ratio	422.14	0.000	302.79	0.000
env_score	21.74	0.000	5.52	0.019
env_ratio	71.16	0.000	16.43	0.000
social_score	67.29	0.000	14.58	0.000
social_ratio	143.88	0.000	25.12	0.000
governance_score	52.76	0.000	0.29	0.588
governance_ratio	58.89	0.000	5.38	0.020

Τα αποτελέσματα του ελέγχου Kruskal-Wallis (Πίνακας 4.4) έδειξαν ότι για την ταξινόμηση ως προς τον κίνδυνο, όλες οι μεταβλητές είναι στατιστικά σημαντικές. Αντίθετα ο έλεγχος ANOVA έδειξε ότι για επίπεδο σημαντικότητας 10% οι μεταβλητές «turnover», «HHI», «cat\_PE», «cat\_EarningsGrowth», «top10assets», «ret3years» και «governance\_score» δεν είναι στατιστικά σημαντικές.

Ο πίνακας συσχετίσεων των μεταβλητών φαίνεται στον Πίνακα 4.5:



Συνοψίζοντας τα αποτελέσματα όλων των προηγούμενων ελέγχων για την επιλογή των μεταβλητών, στον Πίνακα 4.6 παρουσιάζονται συγκεντρωτικά τα συμπεράσματα για τη σημαντικότητα των μεταβλητών για τις τρεις ταξινομήσεις των κεφαλαίων από την Morningstar.

Χρησιμοποιήθηκαν και οι δύο έλεγχοι (ANOVA και Kruskal- Wallis) για να εξαχθούν πιο ολοκληρωμένα συμπεράσματα για τη σημαντικότητα των μεταβλητών εξετάζοντας τόσο την περίπτωση της κανονικής όσο και της μη κανονικής κατανομής για να μειωθούν όσο περισσότερο γίνεται οι μεταβλητές. Στην στήλη «Βάσει KW-ANOVA» είναι τα συμπεράσματα σημαντικότητας που εξήχθησαν από τους δύο ελέγχους, στην στήλη «Βάσει συσχετίσεων» παρουσιάζεται ποιες μεταβλητές έχουν υψηλή συσχέτιση με τις εξαρτημένες μεταβλητές μας βάσει του πίνακα συσχετίσεων και στην στήλη «Τελική επιλογή» παρουσιάζεται ποιες μεταβλητές χρησιμοποιήθηκαν ως στατιστικά σημαντικές και υψηλά συσχετισμένες με τις εξαρτημένες μεταβλητές.

Πίνακας 4.6: Τελική επιλογή μεταβλητών

		Βάσει KW-ANOVA	Βάσει συσχετίσεων	Τελική επιλογή
fund_symbol	fund_symbol			
fund_long_name	fund_long_name			
fund_category	fund_category			
overall_rating	morningstar_overall_rating	DEPENDENT	DEPENDENT	DEPENDENT
risk_rating	morningstar_risk_rating	DEPENDENT	DEPENDENT	DEPENDENT
turnover	annual_holdings_turnover	NAI	NAI	NAI
investment_type	investment_type	NAI	NAI	NAI
size_type	size_type	NAI	NAI	NAI
ER	fund_annual_report_net_expense_ratio	NAI	NAI	NAI
cat_ER	category_annual_report_net_expense_ratio	OXI	OXI	OXI
HHI	Herfindahl- Hirschman Index (HHI)	NAI	NAI	NAI
HHI_sector	ΔΕΙΚΤΗΣ ΔΙΑΦΟΡΟΠΟΙΗΣΗΣ ΑΝΑ ΚΛΑΔΟ	NAI	NAI	NAI
BR	fund_price_book_ratio	NAI	OXI	OXI
cat_BR	Book ratio_fund-categ differ	NAI	OXI	OXI
CR	fund_price_cashflow_ratio	NAI	OXI	OXI
cat_CR	Cashflow Ratio_fund- categ differ	NAI	OXI	OXI
PE	fund_price_earning_ratio	NAI	NAI	NAI
cat_PE	Eatning Ratio_fund- categ differ	NAI	NAI	NAI
PS	fund_price_sales_ratio	NAI	OXI	OXI
cat_PS	Sales Ratio_fund-categ differ	NAI	OXI	OXI
CAP	fund_median_market_cap	NAI	NAI	NAI
cat_CAP	Market Cap_fund-categ differ	NAI	NAI	NAI
EarningsGrowth	fund_year3_earnings_growth	NAI	NAI	NAI
cat_EarningsGrowth	Earnings Growth_fund-categ differ	NAI	NAI	NAI
top10assets	top10_holdings_total_assets	OXI	OXI	OXI
return_rating	morningstar_return_rating	DEPENDENT	DEPENDENT	DEPENDENT
ret1year	ΑΠΟΔΟΣΗ 1 ΕΤΟΥΣ (FUNDS)	NAI	OXI	OXI
ret1year_cat	ΑΠΟΔΟΣΗ 1 ΕΤΟΥΣ (CATEGORY)			
cat_ret1year	ΔΙΑΦΟΡΑ ΑΠΟΔΟΣΗΣ_FUND-CATEG 1 YEAR	NAI	NAI	NAI
ret3years	ΑΠΟΔΟΣΗ 3 ΕΤΩΝ(FUNDS)	NAI	OXI	OXI
ret3years_cat	ΑΠΟΔΟΣΗ 3 ΕΤΩΝ (CATEGORY)			
cat_ret3years	ΔΙΑΦΟΡΑ ΑΠΟΔΟΣΗΣ_FUND-CATEG 3 YEARS	NAI	NAI	NAI
ret5years	ΑΠΟΔΟΣΗ 5 ΕΤΩΝ (FUNDS)	NAI	OXI	OXI
ret5years_cat	ΑΠΟΔΟΣΗ 5 ΕΤΩΝ (CATEGORY)			
cat_ret5years	ΔΙΑΦΟΡΑ ΑΠΟΔΟΣΗΣ_FUND-CATEG 5 YEARS	NAI	NAI	NAI
esg_score	esg_score	OXI	OXI	OXI
peer_esg	peer_esg_avg	OXI	OXI	OXI
esg_ratio	ESG- INDICATOR	OXI	OXI	OXI
env_score	environment_score	NAI	NAI	NAI
peer_environment_avg	peer_environment_avg	OXI	OXI	OXI
env_ratio	ENVIRONMENT- INDICATOR	OXI	OXI	OXI
social_score	social_score	NAI	NAI	NAI
peer_social_avg	peer_social_avg	OXI	OXI	OXI
social_ratio	SOCIAL- INDICATOR	OXI	OXI	OXI
governance_score	governance_score	NAI	NAI	NAI
peer_governance_avg	peer_governance_avg	OXI	OXI	OXI
governance_ratio	GOVERN- INDICATOR	OXI	OXI	OXI

## 4.2 Αποτελέσματα και σχολιασμός

Εφόσον αξιολογήθηκαν τα δεδομένα, έγινε καθαρισμός, συγχωνεύσεις μεταβλητών των δεδομένων και τελική επιλογή των μεταβλητών, προχωρήσαμε στο επόμενο στάδιο της εργασίας εφαρμόζοντας τις εξής τέσσερις μεθόδους μηχανικής μάθησης: Ordinal Logistic Regression, XGBoost, Random Forest και Ordinal Random Forest, μέθοδοι που έχουν αναφερθεί στο προηγούμενο κεφάλαιο της εργασίας. Η υλοποίηση όλων των μεθόδων έγινε σε περιβάλλον R μέσω του πακέτου Caret.

Για την εξαγωγή αξιόπιστων συμπερασμάτων για την διακριτική ικανότητα των παραπάνω μεθόδων, χρησιμοποιήθηκε η διαδικασία 10-fold cross-validation. Αρχικά θα αναλυθεί η σημαντικότητα των μεταβλητών σύμφωνα με τα αποτελέσματα των μεθόδων ταξινόμησης και στην συνέχεια θα γίνει η σύγκριση των μεθόδων ως προς την ακρίβεια της ταξινόμησης.

Παρακάτω παρουσιάζονται το σημαντική είναι ή πόσο επηρεάζει κάθε μεταβλητή τον κάθε έναν δείκτη καθώς και η ακρίβεια που μας δίνει η κάθε μέθοδος:

### 4.2.1 Σημαντικότητα μεταβλητών

Τα αποτελέσματα για την σημαντικότητα των μεταβλητών χρησιμοποιώντας την μέθοδο της μονότονης λογιστικής παλινδρόμησης φαίνονται στον παρακάτω πίνακα 4.7:

Πίνακας 4.7: Σημαντικότητα μεταβλητών μονότονης λογιστικής παλινδρόμησης

	OVERALL RATING	RISK RATING	RETURN RATING
<b>turnover</b>	-0.0400 (10)	0.0034 (0)	-0.0155 (0)
<b>investment_typeGrowth</b>	0.2075 (10)	-0.0135 (0)	0.3144 (10)
<b>investment_typeValue</b>	-0.2883 (10)	0.2679 (10)	-0.1748 (10)
<b>size_typeMedium</b>	0.0841 (0)	-0.0875 (4)	0.0049 (0)
<b>size_typeSmall</b>	0.7247 (10)	-0.4044 (10)	0.3922 (10)
<b>HHI_sector</b>	0.9874 (10)	0.0549 (0)	0.8373 (10)
<b>PE</b>	-0.0898 (10)	0.0092 (3)	-0.0781 (10)
<b>cat_PE</b>	0.1477 (10)	-0.0057 (2)	0.1371 (10)
<b>CAP</b>	0.0171 (10)	-0.0224 (10)	0.0065 (0)
<b>cat_CAP</b>	0.0626 (10)	-0.0291 (10)	0.0445 (10)
<b>EarningsGrowth</b>	0.0020 (0)	0.0119 (10)	0.0041 (0)
<b>cat_EarningsGrowth</b>	0.0364 (10)	-0.0170 (10)	0.0253 (10)
<b>env_score</b>	-0.0187 (1)	0.0004 (0)	-0.0236 (2)



<b>social_score</b>	0.0036 (0)	-0.1621 (10)	-0.1067 (10)
<b>governance_score</b>	0.0156 (0)	0.1776 (10)	0.1407 (10)
<b>ER</b>	-114.4921 (10)	12.8476 (10)	-113.3034 (10)
<b>HHI</b>	-0.6217 (10)	-0.3783 (9)	-0.7402 (10)
<b>BR</b>	-0.0467 (6)	0.0599 (10)	-0.0213 (0)
<b>CR</b>	0.1231 (10)	-0.0252 (9)	0.0971 (10)
<b>cat_ret1year</b>		6.9630 (10)	
<b>cat_ret3years</b>		-6.2958 (10)	
<b>cat_ret5years</b>		2.5770 (10)	

Στον πίνακα 4.7 φαίνονται συγκεντρωτικά η σημαντικότητα των μεταβλητών για τις τρεις κατηγορίες ταξινόμησης χρησιμοποιώντας τον έλεγχο της μονότονης λογιστικής παλινδρόμησης. Στην στήλη της κάθε κατηγορίας φαίνονται οι μέσοι όροι των συντελεστών με τα πρόσημά τους (ανάλογα με το αν έχουν θετική ή αρνητική επίδραση στην αξιολόγηση) και μέσα σε παρένθεση αναφέρεται σε πόσα από τα δέκα υποσύνολα (folds) που εξετάστηκαν ήταν στατιστικά σημαντική η εκάστοτε μεταβλητή σε επίπεδο σημαντικότητας 10%. Όπως φαίνεται στον πίνακα 4.7 οι μεταβλητές «cat\_ret1year», «cat\_ret3years» και «cat\_ret5years» χρησιμοποιήθηκαν μόνο στην ταξινόμηση με βάση τον κίνδυνο. Ο λόγος είναι ότι η ταξινομήσεις με βάση την απόδοση και η ολική αξιολόγηση βασίζονται (μεταξύ άλλων) και στις αποδόσεις των αμοιβαίων κεφαλαίων. Οπότε, εξαιρέθηκαν οι εν λόγω τρεις μεταβλητές γιατί αν ενσωματωθούν σε αυτές τις αναλύσεις, τότε αυτές γίνονται σχεδόν προφανείς. Οπότε εξαιρέθηκαν και εξετάστηκε το πως τα άλλα χαρακτηριστικά των αμοιβαίων κεφαλαίων σχετίζονται με την αξιολόγησή τους από τη Morningstar.

Όσον αφορά την σημαντικότητα των μεταβλητών το αποτέλεσμα της λογιστικής παλινδρόμησης για την ταξινόμηση των κεφαλαίων με βάση τον κίνδυνο (Risk Rating) έδειξε ότι οι μεταβλητές «turnover, investment\_typeGrowth, HHI\_sector και env\_score » δεν είναι στατιστικά σημαντικές για κανένα από τα 10 υποσύνολα αμοιβαίων κεφαλαίων. Όλες οι υπόλοιπες μεταβλητές σχεδόν σε όλες τις δοκιμές του αλγορίθμου είναι στατιστικά σημαντικές σε βαθμό εμπιστοσύνης 99%, εκτός της μεταβλητής «cat\_PE, PE, size\_typeMedium» που όπως παρατηρήσαμε είναι στατιστικά σημαντικές σε 2,3 και 4 υποσύνολα αντίστοιχα.

Όσον αφορά την σημαντικότητα των μεταβλητών το αποτέλεσμα της λογιστικής παλινδρόμησης για την ταξινόμηση των αμοιβαίων κεφαλαίων με βάση την απόδοση (Return Rating) έδειξε ότι οι μεταβλητές «turnover, investment\_typeMedium, BR, CAP, και EarningsGrowth» δεν είναι στατιστικά σημαντικές για κανένα από τα 10 υποσύνολα αμοιβαίων κεφαλαίων. Η μεταβλητή env\_score δεν είναι στατιστικά σημαντική στις 8 από τις 10 περιπτώσεις ενώ στις 2 περιπτώσεις είναι στατιστικά σημαντική για βαθμό εμπιστοσύνης 90%. Όλες οι υπόλοιπες μεταβλητές σχεδόν σε όλες τις δοκιμές του αλγορίθμου είναι

στατιστικά σημαντικές σε βαθμό εμπιστοσύνης 99%, εκτός της μεταβλητής «size\_typeSmall» που όπως παρατηρήσαμε σε 2 υποσύνολα αντίστοιχα φαίνεται ότι είναι στατιστικά σημαντική σε βαθμό 95%.

Επιπλέον όσον αφορά την σημαντικότητα των μεταβλητών το αποτέλεσμα της λογιστικής παλινδρόμησης για την συνολική ταξινόμηση των κεφαλαίων (Overall Rating) έδειξε ότι οι μεταβλητές «size\_typeMedium, EarningsGrowth, env\_score, social\_score, governance\_score» δεν είναι στατιστικά σημαντικές για κανένα από τα 10 υποσύνολα αμοιβαίων κεφαλαίων. Όλες οι υπόλοιπες μεταβλητές σχεδόν σε όλες τις δοκιμές του αλγορίθμου είναι στατιστικά σημαντικές σε βαθμό εμπιστοσύνης 99%, εκτός της μεταβλητής «BR» που όπως παρατηρήσαμε σε 4 υποσύνολα φαίνεται ότι δεν είναι στατιστικά σημαντική, σε 4 φαίνεται σημαντική για βαθμό εμπιστοσύνης 90% ενώ μόνο σε 1 είναι στατιστικά σημαντική για βαθμό εμπιστοσύνης 99%.

Στον παρακάτω πίνακα 4.8 παρουσιάζεται η σημαντικότητα των μεταβλητών στις μεθόδους XGBoost (XGB), Random Forest (RF) και Ordinal Random Forest (ORF) και για τις τρεις κατηγορίες ταξινόμησης.

Πίνακας 4.8: Μέση σημαντικότητα των μεταβλητών στις μεθόδους XGBoost, Random Forest και Ordinal Random Forest

	OVERALL RATING			RETURN RATING			RISK RATING		
	XGB	RF	ORF	XGB	RF	ORF	XGB	RF	ORF
BR	44.91	47.40	61.78	44.77	47.06	64.70	68.49	80.37	98.97
CAP	48.24	47.60	49.36	47.85	47.19	56.51	81.49	88.47	89.31
cat_CAP	81.59	60.78	80.92	67.50	56.00	63.58	99.80	100.00	96.17
cat_EarningsGrowth	63.88	53.36	71.19	61.55	52.72	77.04	70.12	81.63	50.21
cat_PE	85.48	61.12	100.00	81.44	59.21	100.00	67.54	78.12	49.63
cat_ret1year	0	0	0	0	0	0	94.07	97.45	77.01
cat_ret3years	0	0	0	0	0	0	53.66	77.65	43.99
cat_ret5years	0	0	0	0	0	0	46.19	73.43	43.27
CR	53.81	51.94	83.79	54.72	50.82	88.36	58.17	77.81	63.94
EarningsGrowth	47.19	46.21	46.75	45.06	47.05	51.60	59.57	78.12	53.94
env_score	44.42	40.90	42.25	44.36	41.41	47.60	56.27	71.51	53.70
ER	100.00	100.00	47.08	100.00	100.00	49.38	20.75	41.63	2.60
governance_score	34.73	39.83	41.73	41.79	41.31	52.35	51.53	70.26	56.24
HHI	64.40	49.72	41.09	64.43	50.53	46.09	79.30	83.95	49.58
HHI_sector	69.14	50.49	44.36	66.23	51.38	52.19	86.63	87.09	53.92
investment_typeGrowth	0.35	1.54	5.48	1.00	1.66	8.87	0.19	2.91	3.07
investment_typeValue	2.19	3.35	10.09	1.90	3.13	11.86	3.16	5.92	7.91
PE	42.93	46.81	59.27	41.32	46.99	66.19	62.60	77.60	64.98
size_typeMedium	0.59	2.33	1.59	0.39	2.11	1.60	0.83	3.31	2.63
size_typeSmall	0.06	0.00	0.00	0.03	0.00	0.00	0.71	0.00	0.00
social_score	37.14	39.21	35.32	35.60	39.68	38.55	57.02	71.99	63.21
turnover	60.08	49.40	43.76	56.36	49.63	43.85	72.41	79.64	41.33

Στις συγκεκριμένες μεθόδους σε αντίθεση με την λογιστική παλινδρόμηση, η σημαντικότητα των μεταβλητών φαίνεται υπό μορφή ποσοστού. Ως σημαντικότητα εδώ ορίζεται το πόσο επηρεάζει την ακρίβεια ενός μοντέλου κάποια διαταραχή σε μια ανεξάρτητη μεταβλητή. Στη μεταβλητή που επηρεάζει περισσότερο τα αποτελέσματα αποδίδεται σημαντικότητα με το υψηλότερο ποσοστό (100 ή πολύ κοντά στο 100) και όλες οι άλλες ανάγονται σε σχέση με αυτή.

Η μεταβλητή «ER» φαίνεται να είναι η πιο σημαντική στις μεθόδους XGBoost και Random Forest για την συνολική ταξινόμηση και την ταξινόμηση με βάση την απόδοση των κεφαλαίων ενώ φαίνεται να μην επηρεάζει σημαντικά την ταξινόμηση με βάση τον κίνδυνο των κεφαλαίων. Η μεταβλητή «cat\_CAP» είναι η πιο σημαντική για την ταξινόμηση των αμοιβαίων κεφαλαίων με βάση τον κίνδυνο για τις μεθόδους XGBoost και Random Forest ενώ φαίνεται να είναι πολύ σημαντική και για την μέθοδο Ordinal Random Forest μαζί με την μεταβλητή BR. Οι μεταβλητές «cat\_EarningsGrowth», «cat\_PE», «BR», «CR», «EarningsGrowth», «env\_score», «HHI», «HHI\_sector» και «PE» φαίνεται να είναι σημαντικές άλλοτε λιγότερο και άλλοτε περισσότερο σε όλες τις μεθόδους για όλου του είδους τις ταξινομήσεις. Αντίθετα οι μεταβλητές «cat\_ret1year», «cat\_ret3years» και «cat\_ret5years» χρησιμοποιήθηκαν (όπως και στην μονότονη λογιστική παλινδρόμηση) και φαίνεται να είναι σημαντικές μόνο για την ταξινόμηση με βάση τον κίνδυνο του κεφαλαίου, ενώ μεταβολές στις μεταβλητές «investment\_typeGrowth», «investment\_typeValue», «size\_typeMedium» και «size\_typeSmall» φαίνεται να μην επηρεάζουν σημαντικά τα αποτελέσματα σε όλους τους ελέγχους και για τις τρεις κατηγορίες ταξινόμησης.

#### 4.2.2 Αποτελέσματα ταξινόμησης

Στην εν λόγω ενότητα παρουσιάζονται η ακρίβεια ταξινόμησης και τα αποτελέσματα της στατιστικής τιμής Kappa για κάθε μέθοδο. Η στατιστική (ή τιμή) Kappa [15] είναι μια μέτρηση που συγκρίνει μια παρατηρούμενη ακρίβεια με μια αναμενόμενη ακρίβεια (τυχαία πιθανότητα). Ο δείκτης kappa χρησιμοποιείται όχι μόνο για την αξιολόγηση ενός μεμονωμένου ταξινομητή, αλλά και για τη σύγκριση διαφορετικών ταξινομητών. Ο δείκτης αυτός λαμβάνει υπόψη την τυχαία πιθανότητα (σύμφωνα με έναν τυχαίο ταξινομητή), που γενικά σημαίνει ότι είναι λιγότερο παραπλανητικό από τη χρήση απλής ακρίβειας ως μέτρησης (μια παρατηρούμενη ακρίβεια 80% είναι πολύ λιγότερο εντυπωσιακή με μια αναμενόμενη ακρίβεια 75% έναντι αναμενόμενης ακρίβειας 50%). Ο τύπος υπολογισμού φαίνεται από τον παρακάτω τύπο:

$$Kappa = \frac{\text{Παρατηρούμενη ακρίβεια} - \text{Αναμενόμενη ακρίβεια}}{1 - \text{Αναμενόμενη ακρίβεια}}$$

Τα αποτελέσματα για τον δείκτη Kappa της κάθε μεθόδου για την ταξινόμηση με βάση τον κίνδυνο του αμοιβαίου κεφαλαίου φαίνονται στον Πίνακα 4.9.

Πίνακας 4.9: Αποτελέσματα ταξινόμησης (δείκτης kappa) για τα μοντέλα ανάλυσης των κεφαλαίων ως προς τον κίνδυνο

	Ordinal LR	XGBoost	Random Forests	Ordinal RF
Fold 1	5.42%	81.33%	82.04%	81.55%
Fold 2	7.34%	83.98%	86.83%	85.47%
Fold 3	5.89%	82.36%	84.26%	85.42%
Fold 4	4.78%	81.92%	82.64%	83.23%
Fold 5	5.03%	82.59%	85.92%	84.60%
Fold 6	6.58%	81.98%	83.56%	83.73%
Fold 7	7.41%	82.26%	84.30%	84.06%
Fold 8	6.85%	83.95%	86.51%	85.25%
Fold 9	5.02%	82.52%	83.66%	84.86%
Fold 10	5.85%	82.99%	85.14%	84.72%
<b>ΜΕΣΟΣ ΟΡΟΣ</b>	<b>6.02%</b>	<b>82.59%</b>	<b>84.48%</b>	<b>84.29%</b>

Τα αποτελέσματα της ακρίβειας της κάθε μεθόδου για την ταξινόμηση με βάση τον κίνδυνο του αμοιβαίου κεφαλαίου φαίνονται στον Πίνακα 4.10.

Πίνακας 4.10: Αποτελέσματα ακρίβειας ταξινόμησης για τα μοντέλα ανάλυσης των κεφαλαίων ως προς τον κίνδυνο

	Ordinal LR	XGBoost	Random Forests	Ordinal RF
Fold 1	38.58%	86.18%	86.72%	86.29%
Fold 2	37.79%	88.11%	90.23%	89.17%
Fold 3	38.68%	86.93%	88.31%	89.16%
Fold 4	38.36%	86.61%	87.14%	87.57%
Fold 5	38.54%	87.15%	89.60%	88.64%
Fold 6	39.36%	86.70%	87.87%	87.98%
Fold 7	39.89%	86.91%	88.40%	88.19%
Fold 8	37.66%	88.09%	90.00%	89.04%
Fold 9	38.09%	87.02%	87.87%	88.72%
Fold 10	38.81%	87.38%	88.97%	88.65%
<b>ΜΕΣΟΣ ΟΡΟΣ</b>	<b>38.58%</b>	<b>87.11%</b>	<b>88.51%</b>	<b>88.34%</b>

Ο πίνακας 4.10 δείχνει την ακρίβεια της κάθε μεθόδου για την ταξινόμηση των αμοιβαίων κεφαλαίων με βάση τον κίνδυνο, τόσο στο κάθε υποσύνολο (folds) όσο και σαν μέσος όρος

των δέκα υποσυνόλων. Όπως παρατηρούμε η λιγότερο ακριβής μέθοδος είναι αυτή της μονότονης λογιστικής παλινδρόμησης, στην οποία παρατηρούμε μέση ακρίβεια 38,58%. Οι υπόλοιπες τρεις μέθοδοι που εφαρμόσαμε παρουσιάζουν εφάμιλλα αποτελέσματα με την μέθοδο Random Forest να παρουσιάζει οριακά μεγαλύτερη μέση ακρίβεια 88,51%. Συμπεραίνουμε λοιπόν ότι για την ταξινόμηση των αμοιβαίων κεφαλαίων με βάση τον κίνδυνο που διατρέχουν τα κεφάλαια, η μέθοδος που μας δίνει αποτελέσματα πιο ακριβή, συγκριτικά με την αξιολόγηση των αμοιβαίων κεφαλαίων από την Morningstar, είναι η μέθοδος Random Forest.

Τα αποτελέσματα για την στατιστική τιμή Καρρα της κάθε μεθόδου για την ταξινόμηση με βάση την απόδοση του αμοιβαίου κεφαλαίου φαίνονται στον Πίνακα 4.11.

Πίνακας 4.11: Αποτελέσματα ταξινόμησης (δείκτης kappa) για τα μοντέλα ανάλυσης των κεφαλαίων με βάση την απόδοση

	<b>Ordinal LR</b>	<b>XGBoost</b>	<b>Random Forests</b>	<b>Ordinal RF</b>
Fold 1	13.85%	64.26%	65.93%	63.91%
Fold 2	13.58%	60.44%	64.09%	63.95%
Fold 3	14.71%	62.45%	65.20%	63.88%
Fold 4	16.61%	62.07%	66.28%	64.32%
Fold 5	11.89%	63.64%	66.48%	65.22%
Fold 6	16.87%	61.37%	65.75%	65.15%
Fold 7	14.29%	62.45%	67.84%	65.19%
Fold 8	11.86%	63.95%	67.15%	66.97%
Fold 9	17.69%	59.73%	63.04%	63.66%
Fold 10	14.34%	60.67%	65.58%	65.97%
<b>ΜΕΣΟΣ ΟΡΟΣ</b>	<b>14.57%</b>	<b>62.10%</b>	<b>65.73%</b>	<b>64.82%</b>

Τα αποτελέσματα της ακρίβειας της κάθε μεθόδου για την ταξινόμηση με βάση την απόδοση του αμοιβαίου κεφαλαίου φαίνονται στον Πίνακα 4.12.

Πίνακας 4.12: Αποτελέσματα ακρίβειας ταξινόμησης για τα μοντέλα ανάλυσης των κεφαλαίων με βάση την απόδοση

	<b>Ordinal LR</b>	<b>XGBoost</b>	<b>Random Forests</b>	<b>Ordinal RF</b>
Fold 1	42.68%	74.10%	75.16%	73.78%
Fold 2	42.13%	71.49%	74.04%	73.83%
Fold 3	43.30%	72.87%	74.79%	73.72%
Fold 4	44.21%	72.69%	75.56%	74.07%
Fold 5	40.74%	73.62%	75.64%	74.47%
Fold 6	44.31%	71.94%	75.13%	74.39%
Fold 7	42.51%	72.90%	76.73%	74.71%
Fold 8	41.02%	73.96%	76.20%	75.66%
Fold 9	44.90%	71.02%	73.25%	73.35%
Fold 10	42.68%	71.34%	74.84%	75.05%
<b>ΜΕΣΟΣ ΟΡΟΣ</b>	<b>42.85%</b>	<b>72.59%</b>	<b>75.13%</b>	<b>74.30%</b>

Ο πίνακας 4.12 μας δείχνει την ακρίβεια της κάθε μεθόδου για την ταξινόμηση των αμοιβαίων κεφαλαίων με βάση την απόδοση, τόσο στο κάθε υποσύνολο (folds) όσο και σαν μέσος όρος των δέκα υποσυνόλων. Όπως παρατηρούμε η λιγότερο ακριβής μέθοδος είναι αυτή της μονότονης λογιστικής παλινδρόμησης, στην οποία παρατηρούμε μέση ακρίβεια 42,85%. Από τις υπόλοιπες τρεις μεθόδους που εφαρμόσαμε υπερτερεί και πάλι η Random Forest με μέση ακρίβεια 75,13% έναντι της XGBoost (72,59%) και της Ordinal Random Forest (74,30%). Συμπεραίνουμε λοιπόν ότι για την ταξινόμηση των αμοιβαίων κεφαλαίων με βάση την απόδοση των κεφαλαίων, η μέθοδος που παρέχει αποτελέσματα πιο ακριβή, συγκριτικά με την αξιολόγηση των αμοιβαίων κεφαλαίων από την Morningstar, είναι η μέθοδος Random Forest.

Τα αποτελέσματα για την στατιστική τιμή Καρρα της κάθε μεθόδου για την ταξινόμηση με βάση την ολική αξιολόγηση του αμοιβαίου κεφαλαίου φαίνονται στον Πίνακα 4.13.

Πίνακας 4.13: Αποτελέσματα ταξινόμησης (δείκτης kappa) για τα μοντέλα ανάλυσης με βάση την ολική αξιολόγηση των κεφαλαίων

	<b>Ordinal LR</b>	<b>XGBoost</b>	<b>Random Forests</b>	<b>Ordinal RF</b>
Fold 1	18.65%	64.19%	64.78%	65.46%
Fold 2	14.78%	62.24%	66.73%	66.85%
Fold 3	14.55%	61.99%	66.52%	65.42%
Fold 4	17.74%	66.20%	65.91%	65.09%
Fold 5	12.16%	63.82%	64.06%	65.28%
Fold 6	15.20%	63.73%	65.79%	64.74%
Fold 7	17.89%	64.75%	67.57%	66.83%
Fold 8	13.93%	63.16%	64.88%	66.28%
Fold 9	14.71%	63.03%	66.53%	65.81%
Fold 10	17.54%	64.59%	66.12%	64.33%
<b>ΜΕΣΟΣ ΟΡΟΣ</b>	<b>15.72%</b>	<b>63.77%</b>	<b>65.89%</b>	<b>65.61%</b>

Τα αποτελέσματα της ακρίβειας της κάθε μεθόδου για την ταξινόμηση με βάση την ολική αξιολόγηση του αμοιβαίου κεφαλαίου φαίνονται στον Πίνακα 4.14.

Πίνακας 4.14: Αποτελέσματα ακρίβειας ταξινόμησης για τα μοντέλα ανάλυσης των κεφαλαίων με βάση την ολική αξιολόγηση

	<b>Ordinal LR</b>	<b>XGBoost</b>	<b>Random Forests</b>	<b>Ordinal RF</b>
Fold 1	45.16%	74.39%	74.60%	75.13%
Fold 2	42.77%	72.77%	75.96%	75.85%
Fold 3	41.81%	72.55%	75.74%	74.79%
Fold 4	44.96%	75.61%	75.40%	74.34%
Fold 5	40.85%	73.94%	74.15%	74.89%
Fold 6	42.78%	73.89%	75.16%	74.52%
Fold 7	44.59%	74.52%	76.43%	75.69%
Fold 8	42.66%	73.62%	74.79%	75.64%
Fold 9	42.98%	73.19%	75.64%	75.00%
Fold 10	44.80%	74.52%	75.58%	74.20%
<b>ΜΕΣΟΣ ΟΡΟΣ</b>	<b>43.34%</b>	<b>73.90%</b>	<b>75.35%</b>	<b>75.01%</b>

Ο πίνακας 4.14 παρουσιάζει την ακρίβεια της κάθε μεθόδου για την ταξινόμηση των αμοιβαίων κεφαλαίων με βάση την ολική αξιολόγηση των αμοιβαίων κεφαλαίων, τόσο στο κάθε υποσύνολο (folds) όσο και σαν μέσος όρος των δέκα υποσυνόλων. Όπως παρατηρούμε η λιγότερο ακριβής μέθοδος είναι αυτή της μονότονης λογιστικής παλινδρόμησης, στην οποία παρατηρούμε μέση ακρίβεια 43,34%. Από τις υπόλοιπες τρεις μεθόδους που εφαρμόσαμε υπερτερεί και πάλι η Random Forest με μέση ακρίβεια 75,35% έναντι της XGBoost (73,9%) και της Ordinal Random Forest (75,01%). Συμπεραίνουμε λοιπόν ότι για την ταξινόμηση των αμοιβαίων κεφαλαίων με βάση την απόδοση των κεφαλαίων, η μέθοδος που μας δίνει αποτελέσματα πιο ακριβή, συγκριτικά με την αξιολόγηση των αμοιβαίων κεφαλαίων από την Morningstar, είναι η μέθοδος Random Forest.

## 5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Το πρόβλημα της αξιολόγησης των αμοιβαίων κεφαλαίων και γενικά των επενδυτικών σχημάτων έχει αναδειχθεί τα τελευταία χρόνια κυρίως στις ΗΠΑ, όπου οι μορφές επενδύσεων, η δομή και η δραστηριότητα των τραπεζικών ιδρυμάτων είναι διαφορετική απ' ότι στην Ελλάδα και την Ευρωπαϊκή Ένωση.

Υπάρχουν αρκετές εταιρείες αξιολόγησης και κατάταξης των αμοιβαίων κεφαλαίων σε διαβαθμισμένες κατηγορίες ανάλογα με τις επιδόσεις τους στις διαστάσεις της απόδοσης και του κινδύνου. Μία από τις πιο αξιόπιστες θεωρείται η εταιρεία Morningstar, οι αξιολογήσεις της οποίας χρησιμοποιήθηκαν ως σημείο αναφοράς στην παρούσα εργασία. Ένα από τα σημαντικότερα θέματα που έχει ανακύψει είναι το κόστος και η ποιότητα των πληροφοριών που λαμβάνουν οι επενδυτές. Στόχος είναι ο εντοπισμός ενός περιορισμένου συνόλου χαρακτηριστικών, τα οποία περιγράφουν επαρκώς τη δραστηριότητα των αμοιβαίων κεφαλαίων και μπορούν με την εφαρμογή της καταλληλότερης μεθόδου τεχνητής νοημοσύνης, να οδηγήσουν στην ανάπτυξη αξιόπιστων μοντέλων αξιολόγησης.

Δεδομένου του πλήθους των διαφορετικών μεθόδων και αλγορίθμων είναι εμφανές ότι απαιτείται μία πιο ολοκληρωμένη ανάλυση. Βάσει της διαπίστωσης αυτής η παρούσα εργασία επικεντρώθηκε σε ένα συνδυασμό στατιστικών ελέγχων για την επιλογή των μεταβλητών και σε τέσσερις συγκεκριμένους αλγορίθμους για την εφαρμογή της αξιολόγησης των αμοιβαίων κεφαλαίων και της σύγκρισης των αποτελεσμάτων σε σχέση με τις αξιολογήσεις της Morningstar.

Στόχος του πειραματικού σκέλους ήταν να φανεί ποια εκ των μεθόδων μας δίνει μεγαλύτερη ακρίβεια, λαμβάνοντας ως σημείο αναφοράς για την κάθε κατηγορία ταξινόμησης την αξιολόγηση της εταιρείας Morningstar. Τα αποτελέσματα όλων των αναλύσεων συνοψίζονται στον Πίνακα 5.1.

Πίνακας 5.1: Συγκεντρωτικά αποτελέσματα ακρίβειας μεθόδων

	OLR	XGB	RF	ORF
Αξιολόγηση κεφαλαίων με βάση τον κίνδυνο	38.58%	87.11%	88.51%	88.34%
Αξιολόγηση κεφαλαίων με βάση την απόδοση	42.85%	72.59%	75.13%	74.30%
Συνολική αξιολόγηση κεφαλαίων	43.34%	73.90%	75.35%	75.01%
<b>ΜΕΣΟΣ ΟΡΟΣ</b>	<b>41.59%</b>	<b>77.87%</b>	<b>79.66%</b>	<b>79.22%</b>



Τα αποτελέσματα της ανάλυσης έδειξαν ότι η λιγότερο ακριβής μέθοδος με μεγάλη διαφορά από τις υπόλοιπες είναι η μέθοδος της μονότονης λογιστικής παλινδρόμησης, όπου για τις τρεις κατηγορίες ταξινόμησης έδωσε μέση ακρίβεια 41,59% που θεωρείται μη ικανοποιητικό. Σε αντίθεση με τις υπόλοιπες τρεις μεθόδους η μικρότερη ακρίβεια της συγκεκριμένης μεθόδου παρουσιάστηκε στην αξιολόγηση με βάση τον κίνδυνο (38,58%) ενώ η μεγαλύτερη ακρίβεια στην συνολική αξιολόγηση των κεφαλαίων (43,34%). Οι τρεις άλλες μέθοδοι που εφαρμόστηκαν παρουσιάζουν εφάμιλλα αποτελέσματα με μία βραχεία υπεροχή της μεθόδου Random Forest. Η μέθοδος αυτή παρουσιάζει μέση ακρίβεια 79,66% που θεωρείται αρκετά ικανοποιητικό. Η μέγιστη ακρίβεια της μεθόδου παρουσιάστηκε στην αξιολόγηση με βάση τον κίνδυνο (88,51%) ενώ η μικρότερη στην αξιολόγηση με βάση την απόδοση των κεφαλαίων (75,13%). Αντίστοιχα για τις XGBoost και την μονότονη Random Forest οι μέγιστες και ελάχιστες ακρίβειες παρουσιάστηκαν στις ίδιες κατηγορίες αξιολόγησης με μέγιστες 87,11% και 88,34% και ελάχιστες 72,59% και 74,30% αντίστοιχα.

Συνοψίζοντας τα αποτελέσματα της ανάλυσης έδειξαν ότι η μονότονη λογιστική παλινδρόμηση είναι με διαφορά η λιγότερο ακριβής μέθοδος δίνοντας μη ικανοποιητικά αποτελέσματα, ενώ οι υπόλοιπες τρεις μέθοδοι που εφαρμόστηκαν παρέχουν εφάμιλλες ακρίβειες που κρίνονται ικανοποιητικές, με πιο αξιόπιστη και ακριβή την μέθοδο Random Forest.

Οι κύριες προοπτικές και οι αντίστοιχες ερευνητικές κατευθύνσεις εντοπίζονται στην δυνατότητα συνδυασμού των συγκεκριμένων μεθόδων καθώς και στην εξέταση της ακρίβειάς τους σε σχέση και με άλλες αξιόπιστες εταιρείες αξιολόγησης. Επιπλέον θα αποτελούσε ενδιαφέρον αντικείμενο η εξέταση και άλλων αλγορίθμων (πέραν των τεσσάρων της παρούσας εργασίας) ή/ και πιθανοί συνδυασμοί τους, για να διαπιστωθεί η ακρίβειά τους σε σχέση τόσο με τις αξιολογήσεις της Morningstar όσο και με άλλων οίκων αξιολόγησης. Πιθανότατα να μπορεί να εξαχθεί ένα μοντέλο που θα λαμβάνει υπόψιν του ένα περιορισμένο αριθμό μεταβλητών και συνδυάζοντας κάποιες μεθόδους μηχανικής μάθησης, να εξάγει ακριβής αξιολογήσεις συγκριτικά με τις πέντε πιο αναγνωρισμένες εταιρείες αξιολόγησης. Αυτό θα προσέφερε την δυνατότητα, γρήγορων, οικονομικών και αξιόπιστων πληροφοριών για τους επενδυτές.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Πενταράκη Κ. (2003), «Μεθοδολογικό πλαίσιο για την αξιολόγηση της επίδοσης των Αμοιβαίων Κεφαλαίων», Technical University of Crete.,
2. Investment Company FACTBOOK (2022), « A Review of Trends and Activities in the Investment Company Industry», 2022 Investment Company Institute.,
3. O.E.C.D. (2021), « Artificial Intelligence, Machine Learning and Big Data in Finance : Opportunities, Challenges and Implications for Policy Makers»
4. Wang, S. (2003), “Artificial Neural Network”, in *Interdisciplinary Computing in Java Programming, Sringer science and business media, LLC, Triumph Canada*, 81-100.,
5. Krizhevsky A., Sutskever I. and Hinton G. (2017), “ImageNet Classification with Deep Convolutional Neural Networks”, *Communications of the ACM*, Vol. 60/6, 84-90.,
6. Mirete F., Pedro M., Alberto Garcia-Garcia, Juan S., Baixauli-Soler and Maria A. Prats., (2022), « A Review on Machine Learning for Asset Management», *Risks* 10: 84.<https://doi.org/10.3390/risks10040084> Academic Editor: Thorsten Schmidt
7. Gu S., Bryan T. Kelly, Dacheng X., (2021), « Autoencoder asset pricing models», *Journal of Econometrics* 222: 429–50.
8. Arnott, Rob, Campbell R. Harvey και Markowitz H., (2019), «A backtesting protocol in the era of machine learning», *The Journal of Financial Data Science* 1: 64–74.
9. Levenberg K., (1944), «A method for the solution of certain non-linear problems in least squares», *Quarterly of Applied Mathematics* 2: 164–68.
10. Tibshirani R., (1996), «Regression shrinkage and selection via the lasso», *Journal of the Royal Statistical Society: Series B (Methodological)* 58: 267–88
11. Hastie T., Tibshirani R. και Friedman J., (2009), «The elements of statistical learning: data mining, inference, and prediction», *Springer Science & Business Media*.
12. Elith J., Leathwick J.R. και Hastie T., (2008), «A working guide to boosted regression trees», *Journal of Animal Ecology* 77: 802–13
13. Breiman L., (2001) «Random forests», *Machine learning* 45, 1 (2001), 5–32
14. Liaw A., Wiener M., (2002), «Classification and regression by random Forest», *R news* 2, 3 (2002), 18–22.
15. Cohen J., (1960), «A coefficient of agreement for nominal scales», *Educational and Psychological Measurement*. 20 (1): 37–46.
16. Hornung R. (2020), «Ordinal Forests», *Journal of Classification* 37:4–17 (2020)

17. Medeiros M. C., Vasconcelos G.F., Veiga A. και Zilberman E., (2021), «Forecasting inflation in a data-rich environment: the benefits of machine learning methods», *Journal of Business & Economic Statistics* 39 (1), 1-22.
18. Coulombe P. G., Leroux M., Stevanovic D. και Surprenant S., (2020), «How is machine learning useful for macroeconomic forecasting?», *Available in arXiv: <https://arxiv.org/abs/2008.12477>*.
19. DeMiguel V., Gil-Bazo J., Nogales J.F και Santos A.A.P., (2021), «Can Machine Learning Help to Select Portfolios of Mutual Funds?», *Working Papers 1245, Barcelona School of Economics*
20. Zou H. και Hastie T. (2005), «Regularization and variable selection via the elastic net» *Journal of the Royal Statistical Society: Series B* 67 (2), 301- 320.
21. Fafalios S., Charonyktakis P., Tsamardinos I., (2020), «Gradient Boosting Trees», *Gnosis Data Analysis PC*.
22. Breiman L., (1997), «Arcing the edge», Technical Report 486, Statistics Department, University of California.
23. Schapire, R. E. και Freund Y., (2012), «Boosting: Foundations and Algorithms», MIT Press.
24. Tianqi C. και Guestrin C., (2016), «Xgboost: A scalable tree boosting system», *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785-794*.
25. Tianqi C, Guestrin C., (2016), «XGBoost: A Scalable Tree Boosting System», *University of Washington*
26. Wouter V., Baesens B., Bravo C., (2018), « Profit- Driven Business Analytics: A Practitioner's Guide to Transforming Big Data into Added Value», *John Wiley & Sons. Inc, Hoboken.*,
27. Tolles J., και Meurer W.J., (2016), «Logistic Regression Relating Patient Characteristics to Outcomes», *JAMA*. 316 (5): 533–4.
28. Agresti A., (2002), «Categorical Data Analysis», *New York: Wiley, 2002*
29. Le Chap T., (1998), «Applied Categorical Data Analysis», *New York: Wiley, 1998*.
30. Doumpos M., και Zopounidis C., (2002), « Multicriteria Decision Aid Classification Methods», *New York: Springer.*,
31. Rosenblatt F., (1958), «The perceptron: A probabilistic model for information storage and organization in the brain», *Psychological Review* 65: 386 (1958)
32. LeCun Y., Bottou L., Bengio Y., και Patrick H., (1998), «Gradient-based learning applied to document recognition», *Proceedings of the IEEE* 86: 2278–324.

33. Kyunghyun C., Merriënboer B.V, Gulcehre C., Bahdanau D., Bougares F., Schwenk H., και Bengio Y., (2014), «Learning phrase representations using rnn encoder-decoder for statistical machine translation».
34. Hochreiter S., και Schmidhuber J., (1997), «Lstm can solve hard long time lag problems. *Advances in Neural Information Processing Systems*, 9: 473–479.
35. Lloyd S., (1982), «Least squares quantization in pcm», *IEEE Transactions on Information Theory* 28: 129–37
36. Pearson K., (1901), «Liii. on lines and planes of closest fit to systems of points in space», *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2: 559–72.
37. Creswell A., White T., Dumoulin V., Arulkumara B.S., και Bharath A.A., (2018), «Generative adversarial networks: An overview». *IEEE Signal Processing Magazine* 35: 53–65.
38. Sutton R.S., και Barto A.G., (2018). « Reinforcement Learning: An Introduction», *Cambridge: MIT Press*.
39. Watkins C.J. C. H., και Dayan P., (1992), « Q-learning», *Machine Learning*, 8: 279–92.
40. Hester T., Vecerik M., Pietquin O., Lanctot M., Schaul T., Piot B., Horgan D., Quan J., Sendonaris A., Osband I., and et al. (2018), «Deep q-learning from demonstrations», *Paper presented at AAAI Conference on Artificial Intelligence, New Orleans, IL, USA, February 2–7, vol. 32*.
41. Li X., Li L., Gao J., He X., Chen J., Deng L., και He J.. (2015), « Recurrent reinforcement learning: A hybrid approach», *arXiv arXiv:1509.03044*