



**UNIVERSITY OF THESSALY**

**POLYTECHNICAL SCHOOL OF**

**ELECTRICAL AND COMPUTER ENGINEERING**

Εξόρυξη Δεδομένων 2024–2025

**Πρόβλεψη ανοχής ρίσκου και πρόταση  
ολοκληρωμένου χαρτοφυλακίου**

Πολύκαρπος Καλατζίδης, 3000, pkalatzidis@uth.gr  
Παναγιώτης Γερακούδης, 3051, pgerakoudis@uth.gr

# 1 Εισαγωγή

Ένα από τα πρωταρχικά στάδια στην συνεργασία ενός χρηματοοικονομικού σύμβουλου και του πελάτη του είναι η σωστή δημιουργία του προφίλ ρίσκου του πελάτη. Η διαδικασία αυτή κρίνεται αναγκαία και μείζονος σημασίας καθώς καθορίζει σε μεγάλο βαθμό τις αποφάσεις που θα πάρει μετέπειτα ο επαγγελματίας όταν κριθεί να διαχειριστεί τις επενδύσεις του πελάτη του. Μέχρι πρότινος, η διαδικασία αυτή ήταν αρκετά απλοϊκή και συχνά επέφερε λάθος πρόβλεψη, η οποία επέφερε μετέπειτα και λάθος διαχείριση. Πολλές στατιστικές έρευνες έδειξαν ότι οι περισσότεροι χρηματοοικονομικοί σύμβουλοι τείνουν να πιστεύουν ότι οι πελάτες τους έχουν αρκετά μεγαλύτερη ανοχή σε ρίσκο από ότι εν τέλει πραγματικά έχουν.

Η παρούσα εγγρασία επικεντρώνεται στην ανάπτυξη ενός μοντέλου μηχανικής μάθησης (regression model) για την πρόβλεψη της οικονομικής ανοχής σε ρίσκο (risk tolerance) με βάση ένα σύνολο δημογραφικών και οικονομικών χαρακτηριστικών. Η ανοχή στον "κίνδυνο" αποτελεί θεμελιώδη παράγοντα στη διαχείριση προσωπικών οικονομικών, καθώς επηρεάζει τις επενδυτικές αποφάσεις, τις συνήθειες αποταμίευσης και τον συνολικό χρηματοοικονομικό προγραμματισμό. Ένα αξιόπιστο μοντέλο πρόβλεψης επιτρέπει την παροχή εξατομικευμένων οικονομικών συμβουλών, ενώ προσφέρει και βαθύτερη κατανόηση στη συμπεριφορική χρηματοοικονομική. Έπειτα, δημιουργείται ένα ερωτηματολόγιο βασικών ερωτήσεων το οποίο θα χρησιμοποιείται από το μοντέλο μηχανικής μάθησης ώστε να γίνει η πρόβλεψη του ρίσκου που μπορεί να ανεχτεί ένας χρήστης. Τέλος, γίνεται μια πρόταση ενός ολοκληρωμένου χαρτοφυλακίου στηριζόμενη στην ανοχή σε ρίσκο του χρήστη και τις αρχές της θεωρίας Modern Portfolio Theory.

## 2 Σύνολο Δεδομένων

### 2.1 Περιγραφή Συνόλου Δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε σε αυτή τη μελέτη αποτελείται από 23.000 παρατηρήσεις(samples) και 357 χαρακτηριστικά(features). Χρησιμοποιήθηκαν δεδομένα από τη **Federal Reserve Board – Survey of Consumer Finances (SCF), 2022**, που αποτυπώνουν δημογραφικά και οικονομικά χαρακτηριστικά νοικοκυριών στις Η.Π.Α. με στόχο την κατανόηση των επενδυτικών συνθηκών, της οικονομικής γνώσης και των προτιμήσεων απέναντι στον κίνδυνο στον γενικό πληθυσμό καθώς το SCF είναι ένα διαχρονικό, εθνικά αντιπροσωπευτικό δείγμα αμερικανικών νοικοκυριών που καλύπτει αναλυτικά στοιχεία για εισόδημα, περιουσία, χρέη, αποταμίευση και επενδυτικές συμπεριφορές.

### 2.2 Επιλογή Χαρακτηριστικών

Για τη συγκεκριμένη ανάλυση, δόθηκε ιδιαίτερη έμφαση σε δημογραφικά, οικονομικά και ψυχομετρικά χαρακτηριστικά, τα οποία επηρεάζουν σημαντικά την επενδυτική συμπεριφορά και τις αποφάσεις κινδύνου. Από το πλήρες σύνολο δεδομένων επιλέχθηκαν 19 μεταβλητές που θεωρήθηκαν σχετικές και επεξηγηματικές για την πρόβλεψη της ανοχής στον οικονομικό κίνδυνο (risk tolerance). Η επιλογή έγινε με βάση **βιβλιογραφικές πηγές, στατιστική σημασία και πρακτική ερμηνευσιμότητα**.

Τα επιλεγμένα χαρακτηριστικά καλύπτουν:

- **Δημογραφικά:** ηλικία (AGE), φύλο (RACECL), οικογενειακή κατάσταση (MARRIED), αριθμός παιδιών (KIDS), εκπαίδευση (EDUC).
- **Οικονομικά:** εισόδημα (INCOME), καθαρή περιουσία (NETWORTH), αναλογία χρέους προς εισόδημα (DEBT2INC), λόγος μόχλευσης (LEVRATIO).

- **Επενδυτική συμπεριφορά:** κατοχή μετοχών (STOCKS), ομόλογα (BOND, HBOND), ρευστότητα συνταξιοδοτικών επενδύσεων (RETQLIQ), αριθμός μετοχικών συμμετοχών (NSTOCKS).
- **Συνήθειες και στάσεις:** αποταμίευση (WSAVED), τάση για υπερκατανάλωση (SPENDMOR), αποφυγή εξόδων (SPENDLESS), οικονομική γνώση (KNOWL), αποταμιευτική ετοιμότητα σε έκτακτα έξοδα (EMERGSAV).

Αυτός ο συνδυασμός επιτρέπει μια ολιστική κατανόηση των οικονομικών αποφάσεων του κάθε ατόμου.

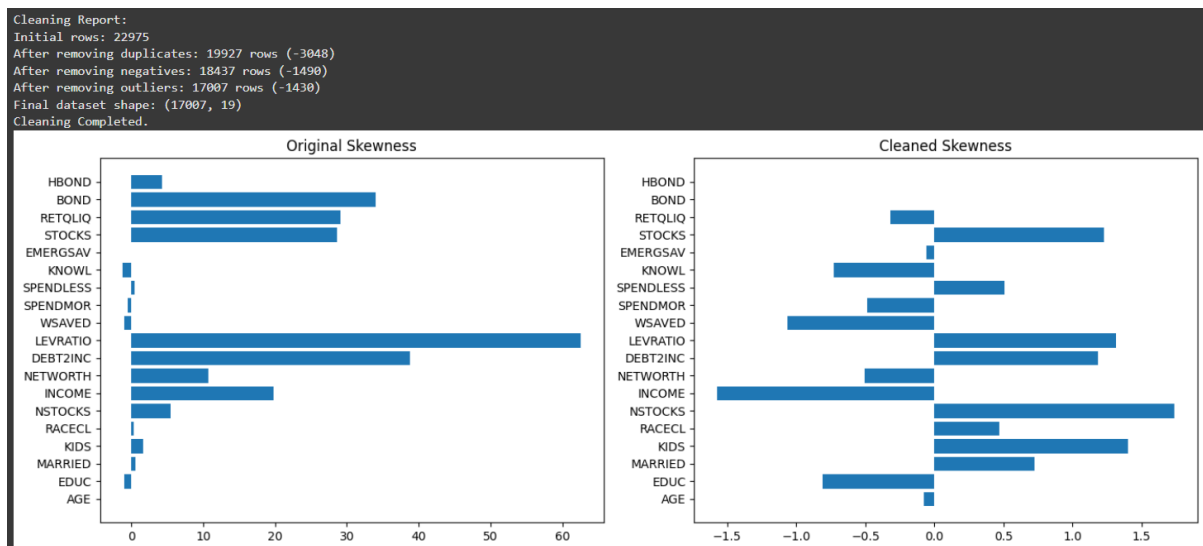
## 2.3 Προεπεξεργασία Δεδομένων

Στο στάδιο της προεπεξεργασίας των δεδομένων, πραγματοποιήθηκε εκτενής καθαρισμός του συνόλου δεδομένων με σκοπό τη διασφάλιση της ποιότητας και της αξιοπιστίας των αποτελεσμάτων.

Ακολούθησαν τα εξής βήματα:

- **Διαγραφή μη ρεαλιστικών ή ασύμβατων τιμών** (π.χ. αρνητικά εισοδήματα, περιουσίες, ή αριθμοί μετοχών).
- **Αφαίρεση διπλότυπων εγγραφών μέσω drop duplicates()**, που διατήρησε μόνο τις μοναδικές περιπτώσεις.
- **Ανίχνευση και αντιμετώπιση ακραίων τιμών (outliers)**
- **Ανάλυση ασυμμετρίας (skewness)** σε συνεχείς μεταβλητές και εφαρμογή λογαριθμικών μετασχηματισμών ( $\ln$ ,  $\log$ ) όπου κρίθηκε απαραίτητο, για σταθεροποίηση διασποράς.
- **Ελεγχος συσχέτισεων** με correlation matrix για την αποφυγή πολυσυγγραμικότητας.

Το τελικό σύνολο δεδομένων διαμορφώθηκε ώστε να είναι καθαρό, ισορροπημένο και κατάλληλο για την επόμενη φάση της μοντελοποίησης.



Σχήμα 1: Data cleaning report

### 3 Δείκτης Ανοχής στον Κίνδυνο

#### 3.1 Υπολογισμός Δείκτη Ανοχής στον Κίνδυνο (Risk Tolerance)

Για τις ανάγκες της μελέτης, ορίστηκε ένας συνθετικός δείκτης "Risk Tolerance" με τιμές μεταξύ 0 και 1, που εκτιμά το επίπεδο οικονομικής ανοχής στον επενδυτικό κίνδυνο ενός ατόμου. Ο υπολογισμός του αποτελεί την κεντρική μεταβλητή στόχο (target variable) του προβλήματος πρόβλεψης. Η μεταβλητή αυτή δεν υπήρχε αυτούσια στο dataset ο τύπος υπολογισμού στηρίζεται σε θεωρητικά και εμπειρικά μοντέλα από τη χρηματοοικονομική θεωρία. Ο υπολογισμός περιλαμβάνει λογαριθμική κανονικοποίηση επιλεγμένων μεταβλητών όπως η κατοχή μετοχών, αριθμός επενδυτικών προϊόντων, ρευστά διαθέσιμα για συνταξιοδότηση, καθώς και δείκτες χρέους.

Αναλυτικότερα:

```
import seaborn as sns

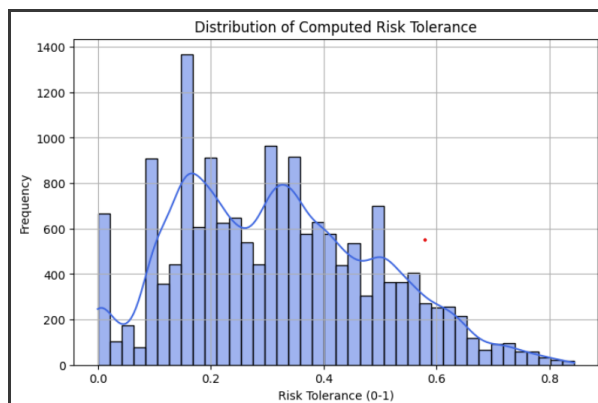
def norm_log(x, divider=5):
    return np.log1p(max(x, 0)) / divider

def compute_risk_tolerance(row):
    score = (
        0.25 * norm_log(row['STOCKS']) +
        0.15 * norm_log(row['NSTOCKS']) +
        0.20 * norm_log(row['BOND']) +
        0.10 * int(row['HBOND'] > 0) +
        0.10 * norm_log(row['RETQLIQ']) +
        0.10 * row['SPENDMOR'] +
        0.10 * (1 - row['SPENDLESS']) +
        0.05 * row['DEBT2INC'] +
        0.05 * row['LEVRATIO']
    )
    return round(min(max(score, 0), 1.0), 4)
```

Σχήμα 2: Risk Tolerance Computation

#### 3.2 Κατανομή Ανοχής στον Κίνδυνο

Το γράφημα κατανομής παρουσιάζει μια ασύμμετρη δεξιά κατανομή με μέγιστη συγκέντρωση γύρω από τις τιμές 0.2–0.4. Αυτό υποδηλώνει ότι το μεγαλύτερο μέρος του πληθυσμού εκφράζει μέτρια ή χαμηλή ανοχή στον επενδυτικό κίνδυνο. Η κατανομή είναι συνεχής με παρουσία αρκετών τοπικών κορυφών, ένδειξη της ετερογένειας στη χρηματοοικονομική συμπεριφορά των νοικοκυριών και ελάχιστοι συμμετέχοντες εμφανίζουν πολύ υψηλή ανοχή (> 0.7).



Σχήμα 3: Distribution of Risk Tolerance

## 4 Μοντέλο Πρόβλεψης Ρίσκου

### 4.1 Επιλογή Μεθόδου

Ως στόχος τέθηκε η πρόβλεψη της υπολογιζόμενης μεταβλητής Risk Tolerance μέσω εποπτευόμενης μάθησης (supervised learning) και αλγόριθμοι παλινδρόμησης (regression models), κατάλληλοι για συνεχή μεταβλητή στόχο. Αρχικά, είναι σημαντικό να σημειωθεί ότι οι μεταβλητές που χρησιμοποιήθηκαν για τον υπολογισμό του δείκτη Risk Tolerance έχουν απορριφθεί και τα μοντέλα εκπαιδεύονται πάνω στις υπόλοιπες επιλεγμένες μεταβλητές.

Εξετάστηκαν δέκα διαφορετικοί αλγόριθμοι:

- **Γραμμικοί:** Linear, Ridge, Lasso Regression
- **Μη-Γραμμικοί:** SVR, KNN, Decision Tree
- **Ensemble methods:** Random Forest, XGBoost, Gradient Boosting
- **Νευρωνικά Δίκτυα:** Multi-Layer Perceptron

### 4.2 Εκπαίδευση και Αξιολόγηση

Το σύνολο δεδομένων χωρίστηκε σε σύνολο εκπαίδευσης (80%) και ελέγχου (20%). Πραγματοποιήθηκε scaling όπου κρίθηκε απαραίτητο (κυρίως για γραμμικά μοντέλα), ενώ για τα δενδροειδή μοντέλα δεν ήταν αναγκαία η κανονικοποίηση. Έγινε **cross-validation 5-fold** για αξιολόγηση γενίκευσης και αποφυγή overfitting.

Τα μοντέλα αξιολογήθηκαν με βάση τριών μετρικών:

- **Root Mean Squared Error (RMSE):** Μέση τετραγωνική απόκλιση μεταξύ πραγματικών και προβλεπόμενων τιμών.
- **Mean Absolute Error (MAE):** Αντιπροσωπεύει το μέσο μέγεθος σφάλματος πρόβλεψης.
- **Coefficient of Determination ( $R^2$ ):** Αντικατοπτρίζει το ποσοστό της διακύμανσης του στόχου που εξηγείται από το μοντέλο.

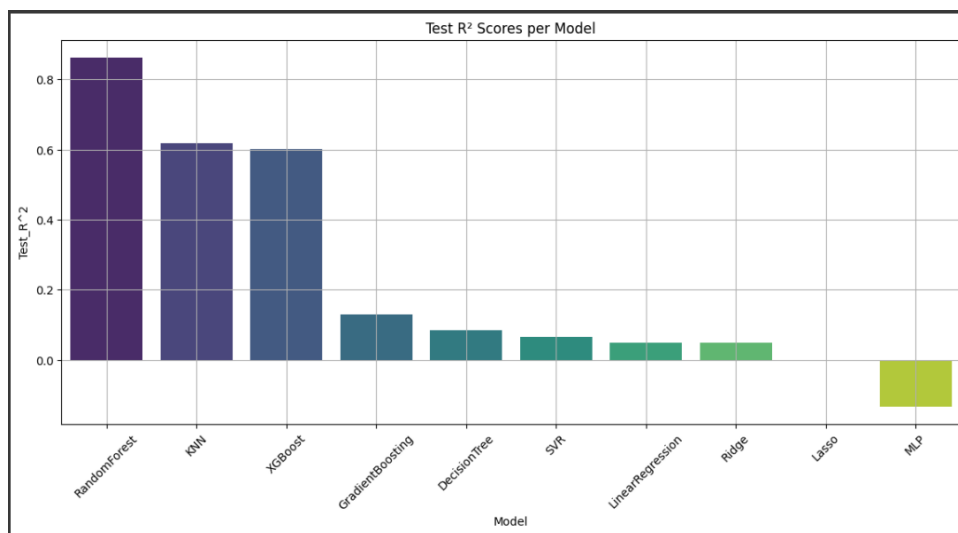
| DESCENDING COMPARISON SUMMARY |                  |             |            |           |           |          |
|-------------------------------|------------------|-------------|------------|-----------|-----------|----------|
|                               | Model            | CV_Mean_R^2 | CV_Std_R^2 | Test_R^2  | Test_RMSE | Test_MAE |
| 5                             | RandomForest     | 0.799646    | 0.009760   | 0.862477  | 0.066275  | 0.040259 |
| 3                             | KNN              | 0.453992    | 0.016258   | 0.618442  | 0.110393  | 0.079941 |
| 9                             | XGBoost          | 0.566452    | 0.006668   | 0.601691  | 0.112790  | 0.085968 |
| 6                             | GradientBoosting | 0.118861    | 0.004047   | 0.130613  | 0.166635  | 0.137221 |
| 4                             | DecisionTree     | 0.070207    | 0.006449   | 0.084573  | 0.170990  | 0.140460 |
| 7                             | SVR              | 0.047226    | 0.005941   | 0.065193  | 0.172791  | 0.141791 |
| 0                             | LinearRegression | 0.039624    | 0.004666   | 0.048567  | 0.174321  | 0.143617 |
| 1                             | Ridge            | 0.039624    | 0.004665   | 0.048567  | 0.174321  | 0.143617 |
| 2                             | Lasso            | -0.001021   | 0.001311   | -0.000059 | 0.178720  | 0.147366 |
| 8                             | MLP              | 0.001649    | 0.026991   | -0.133297 | 0.190253  | 0.160975 |

Σχήμα 4: Comparison(CV & Test) Summary 10 model

Κύρια μετρική: **R<sup>2</sup> score στο σύνολο ελέγχου**

Σύμφωνα με το παρακάτω γράφημα:

- Το **Random Forest** εμφάνισε την υψηλότερη επίδοση ( $R^2 \approx 0.87$ ).
- Τα μοντέλα **KNN** και **XGBoost** επίσης σημείωσαν υψηλές επιδόσεις ( $R^2 \approx 0.6$ ).
- Τα γραμμικά μοντέλα (Linear, Ridge, Lasso) είχαν χαμηλές επιδόσεις ( $R^2 \approx 0.05$ ).
- Το **MLP Regressor** παρουσίασε αρνητική επίδοση ( $R^2 < 0$ )



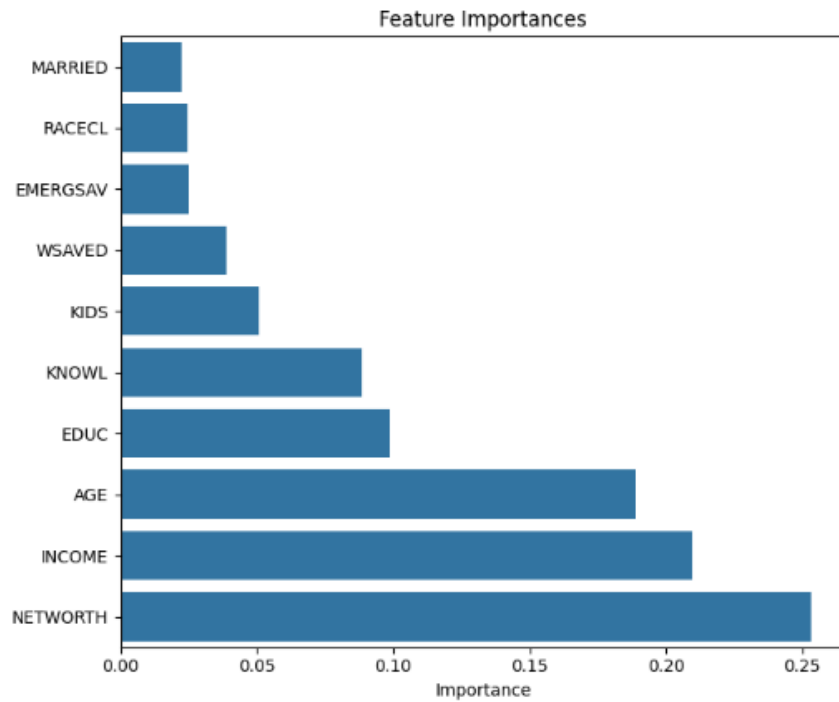
Σχήμα 5: Barplot 10 Model

### 4.3 Ανάλυση Απόδοσης Βέλτιστου Μοντέλου

Μεταξύ των δέκα υποψήφιων μοντέλων, το **Random Forest** παρουσίασε εξαιρετική ισορροπία μεταξύ προσαρμογής και γενίκευσης. Η μέση απόδοση σε 5-πλή cross-validation έδωσε  $CV R^2 \approx 0.7996$ , ενώ η τελική αξιολόγηση σε ανεξάρτητο test set ανέδειξε ένα  $R^2 \approx 0.8625$ . Το υψηλό  $R^2$  στο test set, συνδυασμένο με το cross-validation score, υποδηλώνει ότι το μοντέλο έχει μάθει ουσιώδη μοτίβα στα δεδομένα χωρίς να τα υπερεκπαίδευσε. Η μικρή διαφορά ( $\approx 6\%$ ) μεταξύ των δύο τιμών δείχνει ότι το μοντέλο αποφεύγει το φαινόμενο της υπερεκπαίδευσης και είναι σε θέση να αποδώσει αξιόπιστες προβλέψεις σε νέα, άγνωστα δεδομένα γεγονός που ενισχύει την εμπιστοσύνη στο τελικό του αποτέλεσμα.

Το Random Forest αποδείχθηκε σταθερό, ανθεκτικό στον θόρυβο και ιδιαίτερα αποδοτικό στην ανάλυση της σχέσης μεταξύ των δημογραφικών και οικονομικών χαρακτηριστικών με το δείκτη ανοχής ρίσκου ο οποίος μπορεί να προβλεφθεί αποτελεσματικά, προσφέροντας χρήσιμες ενδείξεις για εφαρμογές σε χρηματοοικονομικό σχεδιασμό, προσωποποιημένες επενδυτικές στρατηγικές και συμπεριφορική οικονομία.

Επιπλέον η ανάλυση των σημαντικότερων μεταβλητών, όπως φαίνεται και από την παρακάτω εικόνα, βάσει του μοντέλου Random Forest δείχνει ότι οι μεταβλητές (NETWORTH), (INCOME) και (AGE) συνεισφέρουν με διαφορά περισσότερο στην πρόβλεψη του δείκτη ανοχής ρίσκου γεγονός που υποδεικνύει ότι οικονομικοί και δημογραφικοί παράγοντες παίζουν κρίσιμο ρόλο στον προσδιορισμό της ανοχής ρίσκου. Αντίθετα, μεταβλητές (MARRIED) και (RACECL) έχουν ελάχιστη συνεισφορά στην τελική πρόβλεψη. Αυτή η κατανομή ευθυγραμμίζεται με τη χρηματοοικονομική θεωρία, η οποία συχνά συνδέει την καθαρή περιουσία και το εισόδημα με αυξημένη ικανότητα ανάληψης κινδύνου.



Σχήμα 6: Feature Importance RF

## 5 Ερωτηματολόγιο Χρήστη

### 5.1 Δομή και Ερωτήσεις

Το ερωτηματολόγιο που καλείται να απαντήσει ο χρήστης συντάχθηκε βάση έρευνας αντίστοιχων ερωτηματολογίων στον χώρο των επενδύσεων. Κυρίαρχο ρόλο είχαν και οι ερωτήσεις της Έρευνας Οικονομικών Καταστάσεων (Survey of Consumer Finances - SCF 2022) από το συμβούλιο ομοσπονδιακών αποθεματικών των ΗΠΑ. Πιο συγκεκριμένα, οι κατηγορίες των ερωτήσεων και οι αντιστοιχίσεις των απαντήσεων σε μεταβλητές:

Προσωπικές Πληροφορίες:

- Ηλικία - 'AGE'
- Φύλο - 'RACECL'
- Οικογενειακή Κατάσταση - 'MARRIED'
- Αριθμός Παιδιών - 'KIDS'
- Επίπεδο Εκπαίδευσης - 'EDUC'
- Ετήσιο Εισόδημα - 'INCOME'
- Συνολική Καθαρή Αξία - 'NETWORTH'

Οικονομικές Προτιμήσεις και Συμπεριφορές:

- Αν λάβετε απροσδόκητα 1.000 ευρώ, πώς θα τα χρησιμοποιούσατε κυρίως;- 'WSAVED'
- Αν το νοικοκυριό σας είχε αύξηση μόνιμου εισοδήματος, πώς θα αντιδρούσατε σε σχέση με τις δαπάνες; - 'SPENDMOR'
- Αν το νοικοκυριό σας είχε μείωση μόνιμου εισοδήματος, πώς θα αντιδρούσε; - 'SPENDLESS'

- Διαθέτει το νοικοκυριό σας αποταμιεύσεις για έκτακτες ανάγκες; - 'EMERGSAY;
- Πόσο καλά αισθάνεστε ότι κατανοείτε βασικά χρηματοοικονομικά ζητήματα; - 'KNOWL'

Επενδυτικός Ορίζοντας

Σχήμα 7: Questionnaire.1

Σχήμα 8: Questionnaire.2

## 5.2 Χρήση στο Μοντέλο

Το ερωτηματολόγιο χρησιμοποιείται για τη συλλογή δημογραφικών, οικονομικών και επενδυτικών χαρακτηριστικών των χρηστών. Οι απαντήσεις αποθηκεύονται σε μια βάση δεδομένων μέσω του μοντέλου Questionnaire, το οποίο υλοποιείται στο Django. Κάθε εγγραφή είναι μοναδική με βάση το UUID, ενώ χρησιμοποιείται και session\_key για την ταυτοποίηση συνεδριών. Οι τελικές απαντήσεις του χρήστη μετασχηματίζονται σε αριθμητικά χαρακτηριστικά και χρησιμοποιούνται ως είσοδοι στο μοντέλο πρόβλεψης ανοχής ρίσκου.

## 6 Σύσταση Χαρτοφυλακίου

### 6.1 Εφαρμογή Modern Portfolio Theory

Η πρόταση χαρτοφυλακίου βασίζεται στην Modern Portfolio Theory (MPT), η οποία στοχεύει στη βέλτιστη κατανομή κεφαλαίων μεταξύ διαφόρων περιουσιακών στοιχείων ώστε να επιτευχθεί ο καλύτερος δυνατός συνδυασμός απόδοσης και κινδύνου. Χρησιμοποιείται μια εκδοχή της MPT, η οποία λαμβάνει υπόψη δυναμικές μεταβολές στις συσχετίσεις των επενδύσεων, τον δείκτη Sharpe (απόδοση μιας επένδυσης σε σχέση με τον κίνδυνο), καθώς και δείκτες συγκέντρωσης και διαφοροποίησης. Τα διαθέσιμα επενδυτικά οχήματα που συμπεριλήφθηκαν είναι τα εξής:



- US Stocks (VTI) — Μετοχές ΗΠΑ
- Gold (GLD) — Χρυσός
- Real Estate (VNQ) — Ακίνητη Περιουσία
- Corporate Bonds (LQD) — Εταιρικά Ομόλογα
- Treasuries (TLT) — Κρατικά Ομόλογα ΗΠΑ
- Cryptocurrency (BTC-USD) — Κρυπτονόμισμα (Bitcoin)
- Emerging Markets (VWO) — Αναδυόμενες Αγορές
- Small Cap Value (IWN) — Μετοχές Μικρής Κεφαλαιοποίησης
- Cash Equivalents — Ισοδύναμα Μετρητών

## 6.2 Αντιστοίχιση Ρίσκου με Επενδύσεις

Το επίπεδο ανοχής ρίσκου του χρήστη(risk tolerance) καθορίζει τη σύνθεση του χαρτοφυλακίου σε μεγάλο βαθμό καθώς το κυρίαρχο και ζητούμενο μετρικό είναι η μεταβλητότητα του χαρτοφυλακίου . Η στοχευόμενη μεταβλητότητα υπολογίζεται με βάση την παρακάτω εξίσωση:

$$\text{Στοχευόμενη Μεταβλητότητα} = 0.04 + (\text{Risk Tolerance} \times 0.31)$$

Η ελάχιστη στοχευόμενη μεταβλητότητα είναι 4%, και αυξάνεται γραμμικά μέχρι και 35% καθώς αυξάνεται το επίπεδο ανοχής ρίσκου.

Έτσι ανάλογα με την ανοχή ρίσκου του χρήστη στην σύνθεση του χαρτοφυλακίου παρατηρούνται οι εξής διαφοροποιήσεις:

- **Χαμηλό ρίσκο:** αποκλείονται περιουσιακά στοιχεία υψηλής μεταβλητότητας (π.χ. κρυπτονομίσματα), και δίνεται έμφαση σε ομόλογα και μετρητά.
- **Μέτριο ρίσκο:** διατηρείται ισορροπία μεταξύ μετοχών και σταθερού εισοδήματος.
- **Υψηλό ρίσκο:** ενισχύεται η παρουσία αγορών με υψηλές αποδόσεις (π.χ. αναδυόμενες αγορές, κρυπτονομίσματα).

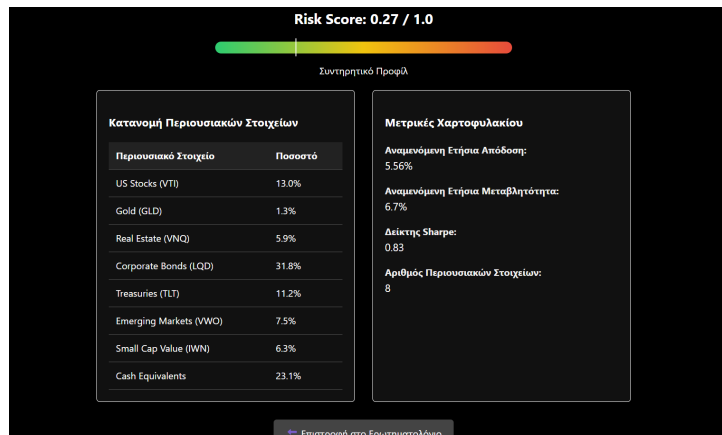
Εφαρμόζονται επίσης δυναμικά όρια ανά κατηγορία επένδυσης, τα οποία εξαρτώνται από το επίπεδο κινδύνου. Οι συσχετίσεις των στοιχείων διαφοροποιούνται ώστε να ενισχύεται η αποδοτικότητα του χαρτοφυλακίου μέσω διαφοροποίησης.

## 6.3 Τελική Παρουσίαση Χαρτοφυλακίου

Η έξοδος της συνάρτησης περιλαμβάνει:

- Το τελικό ποσοστό κατανομής σε κάθε περιουσιακό στοιχείο.
- Εκτιμώμενη απόδοση, μεταβλητότητα και δείκτης Sharpe.
- Κατηγορία επενδυτικού προφίλ (π.χ. Συντηρητικός, Επιθετικός).
- Βαθμός διαφοροποίησης (αριθμός διαφορετικών επενδύσεων).

Η κατανομή περιορίζει την υπερσυγκέντρωση και ενισχύει τη σταθερότητα του χαρτοφυλακίου, σύμφωνα με τις ατομικές προτιμήσεις ρίσκου του χρήστη.



Σχήμα 9: recommended portfolio

## 7 Χρησιμοποιούμενα Εργαλεία

Η ανάπτυξη της εφαρμογής πραγματοποιήθηκε με τη χρήση των παρακάτω εργαλείων:

- **Python 3.10+** – Γλώσσα προγραμματισμού για όλη την ανάλυση δεδομένων, μοντελοποίηση και ανάπτυξη backend.
- **Django** – Web framework για τη δημιουργία του backend και του API της εφαρμογής.
- **SQLite** – Ελαφριά σχεσιακή βάση δεδομένων για την αποθήκευση των απαντήσεων του ερωτηματολογίου.
- **Pandas** – Βιβλιοθήκη για τη διαχείριση και ανάλυση δεδομένων.
- **NumPy** – Υπολογιστική βιβλιοθήκη για διανυσματικές πράξεις και υπολογισμό στατιστικών μετρικών.
- **SciPy** – Χρησιμοποιήθηκε για την επίλυση του προβλήματος βελτιστοποίησης χαρτοφυλακίου μέσω της μεθόδου SLSQP.
- **Matplotlib / Seaborn** – Οπτικοποίηση δεδομένων, προαιρετικά για debugging ή παρουσίαση αποτελεσμάτων.
- **HTML / CSS / JavaScript** – Για τη δημιουργία του frontend της εφαρμογής (φόρμα ερωτηματολογίου και αποτελέσματα).

## Αναφορές

- [1] Federal Reserve Board. *Survey of Consumer Finances (SCF)*, 2022.  
<https://www.federalreserve.gov/econres/scfindex.htm>
- [2] Markowitz, H. (1952). *Portfolio Selection*. The Journal of Finance, 7(1), 77–91.
- [3] Sharpe, W. F. (1966). *Mutual Fund Performance*. Journal of Business, 39(1), 119–138.
- [4] Holzhauser, H. M., Lu, X., McLeod, R., & Wang, J. (2023). *RiskTRACK: The Five-Factor Model for Measuring Risk Tolerance*. Journal of Risk Finance.

- [5] Boyd, S., & Paleyes, A. (2023). *Machine Learning for Science: Blueprints for Finance*. O'Reilly Media.