



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

---

## Τελική Απαλλακτική Εργασία

Φοιτητές: *Ιάκωβος Μαστρογιαννόπουλος - Κωνσταντίνος Καμαρόπουλος*

---

Μάθημα: *Διαχείριση Γνώσης* – Καθηγητής: *Αθανάσιος Κιούρτης*

### Περίληψη

Αυτή είναι η τελική εργασία για το μάθημα «Διαχείριση Γνώσης» των φοιτητών:

- Ιάκωβος Μαστρογιαννόπουλος, 713242017102
- Κωνσταντίνος Καμαρόπουλος, 71346830

## Περιεχόμενα

<b>1</b>	<b>Διαδικασία ανακάλυψης γνώσης</b>	<b>3</b>
1.1	Επιλογή σεναρίου . . . . .	3
1.2	Επιλογή δεδομένων . . . . .	3
1.3	Εξόρυξη δεδομένων . . . . .	4
1.3.1	Υλοποίηση και αποτελέσματα KMeans . . . . .	4
1.3.2	Υλοποίηση και αποτελέσματα MeanShift . . . . .	5
1.3.3	Σχολιασμός αποτελεσμάτων . . . . .	6
<b>2</b>	<b>Θεωρητικό υπόβαθρο της διαχείρισης γνώσης</b>	<b>7</b>
2.1	Διαχωρισμός δεδομένων, πληροφορίας και γνώσης . . . . .	7
2.2	Κύκλος ζωής διαχείρισης γνώσης . . . . .	7
2.3	Μοντέλο αλλαγών του Kotter . . . . .	7

## Κατάλογος σχημάτων

1	Παράδειγμα Clustering . . . . .	4
2	Αποτελέσματα KMeans . . . . .	5
3	Αποτελέσματα MeanShift . . . . .	6

## Κατάλογος πινάκων

## 1. Διαδικασία ανακάλυψης γνώσης

Στο πρώτο τμήμα της εργασίας, θα γίνει η ανάλυση του σεναρίου που επιλέχτηκε και θα παρουσιαστεί το πως πραγματοποιήθηκε η διαχείριση των δεδομένων για να βγει το συμπέρασμα και η γνώση στην τελική της μορφή.

**1.1. Επιλογή σεναρίου.** Σε όποιον αρέσει να ταξιδεύει στον κόσμο, κατά μεγάλη πιθανότητα θα έχει βρεθεί σε μία κατάσταση που δεν θα μπορεί να εύκολα να αποφύγει. Σε αρκετές χώρες, το νόμισμα το οποίο έχουν δεν είναι το νόμισμα της χώρας που κατοικεί και συνήθως αναγκάζετε ο ταξιδιώτης να ανταλλάξει τα χρήματα του στο τοπικό νόμισμα. Ένας τρόπος που μπορεί να αποφύγει, βέβαια το πρόβλημα είναι να χρησιμοποιεί μόνο την τραπεζική του κάρτα για όλες τις συναλλαγές, αλλά αυτό δεν είναι η λύση του προβλήματος, αλλά μόνο η αποφυγή του. Η λύση του προβλήματος αυτού ήταν και η δημιουργία του Ευρώ στα κράτη που ανήκουν Ευρωπαϊκή Ένωση (Ε.Ε.) και στην Ευρωζώνη. Έτσι, ένας Έλληνας μπορεί να ταξιδέψει σε μία χώρα που βρίσκετε μέσα στην Ε.Ε. και με το ίδιο νόμισμα να μπορεί να κάνει συναλλαγές εύκολα. Τι γίνεται όμως στις χώρες που δεν ανήκουν στην Ευρωζώνη;

Αρκετές χώρες έχουν λύση αυτό το πρόβλημα με την γραφείων των συναλλαγών (trade currency offices) που σκοπός τους είναι να πηγαίνει ο ταξιδιώτης και να μπορεί να ανταλλάξει το νόμισμα του στο τοπικό νόμισμα με κέρδος προς το γραφείο. Εκεί όμως ξεκινάει ένα άλλο πρόβλημα, που έχει παρουσιαστεί πολλές φορές σε βίντεο, το πρόβλημα της εκμετάλλευσης. Πολλά από αυτά τα γραφεία, βάζουν υπερβολικά μεγάλα ποσοστά και κρατάνε πολλή μεγάλο χρηματικό πόσο. Αυτό με κάποιον τρόπο θα πρέπει να βρεθεί μία λύση για να σταματήσει να υπάρχει αυτό το πρόβλημα.

**1.2. Επιλογή δεδομένων.** Τα δεδομένα τα οποία επιλέχτηκαν βρίσκονται σε ένα dataset το οποίο ήταν αποθηκευμένο σε CSV αρχείο, το οποίο βρέθηκε στην ιστοσελίδα του Kaggle. Το συγκεκριμένο αρχείο τροποποιήθηκε κατάλληλα και κρατήθηκαν τα δεδομένα τα οποία ήταν χρήσιμα για την συγκεκριμένη περίπτωση. Τα δεδομένα αυτά ήταν το exchange rate του Ευρώ σε άλλα νομίσματα.

Χρησιμοποιήθηκε το Pandas module της Python για να διαβάσει τα δεδομένα και να μετατρέψει τις πληροφορίες που συγκεντρώθηκαν, μετατράπηκαν σε Dataframe και αποθηκεύτηκαν σε XML. Ο κώδικας είναι ο εξής:

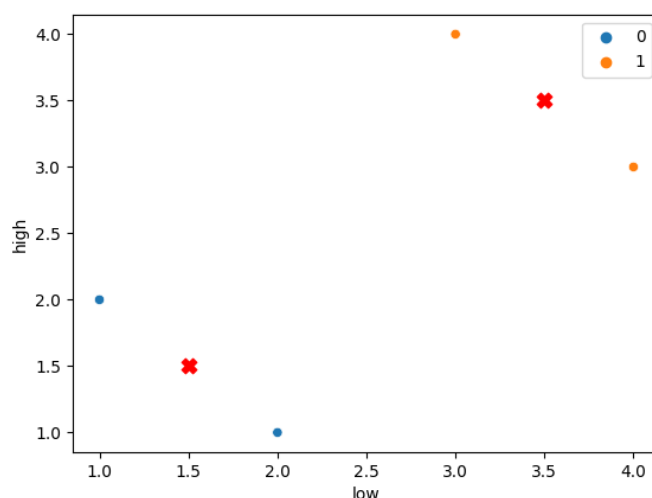
```
# Clean dataset implementation
def create_dataframe(path, dataset_path='./forex.csv') -> pd.DataFrame:
    # Read data from the CSV source file
    dataset = pd.read_csv(dataset_path)
    dataset.to_xml(path)

    # Keep only the info that is needed
    dataset = dataset.where(dataset['currency'] == 'EUR').dropna()
    dataset = dataset.where(dataset['high'] < 2).dropna()
    dataset = dataset.where(dataset['high'] > 0.2).dropna()

    # Dictionary
    clean_dataset = {
        'high': dataset['high'],
        'low': dataset['low'],
        'open': dataset['open'],
        'close': dataset['close']
    }
```

```
# Parse dictionary to new DataFrame, save it and return this back
clean_dataset = pd.DataFrame(clean_dataset)
clean_dataset.to_csv('./dataset.csv')
return clean_dataset
```

**1.3. Εξόρυξη δεδομένων.** Υπάρχουν διάφοροι μέθοδοι που μπορούν να πραγματοποιηθεί εξόρυξη δεδομένων. Η μέθοδος που επιλέχτηκε για την συγκεκριμένη εργασία είναι η συσταδοποίηση (clustering). Ένας clustering αλγόριθμος χωρίζει το δείγμα που δέχεται σε μικρότερες ομάδες με ένα κοινό χαρακτηριστικό. Οι δύο clustering αλγόριθμοι που επιλέχτηκαν είναι ο KMeans και ο MeanShift.

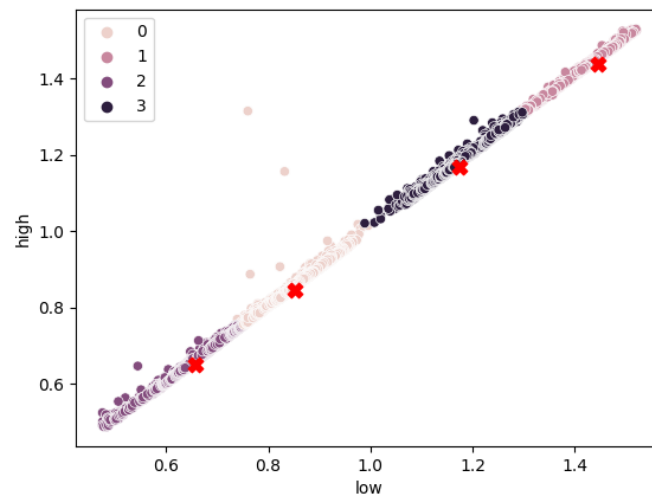


Σχήμα 1: Παράδειγμα Clustering

**1.3.1. Υλοποίηση και αποτελέσματα KMeans.** Ο KMeans είναι ένας απλός clustering αλγόριθμος που βγάζει αρκετά πιστά δεδομένα.

```
# KMeans implementation
def kmeans(
    dataframe: pd.DataFrame, n_clusters=4, path: str = "k_means_cluster.png"
) -> None:
    k_means = KMeans(n_clusters=n_clusters, init="k-means++", random_state=0)
    .fit(dataframe)

    sns.scatterplot(data=dataframe, x="low", y="high", hue=k_means.labels_)
    plt.scatter(
        k_means.cluster_centers_[0], k_means.cluster_centers_[1], marker='X',
        c="r", s=80, label="centroids"
    )
    plt.savefig(path)
    plt.show()
```



Σχήμα 2: Αποτελέσματα KMeans

**1.3.2. Υλοποίηση και αποτελέσματα MeanShift.** Ο MeanShift είναι ένας εναλλακτικός clustering αλγόριθμος ο οποίος είναι λίγο πιο αργός αλλά τείνει στο να βγάζει πιο πιστά αποτελέσματα από ότι ο KMeans.

```
# MeanShift implementation
```

```
def mean_shift(dataframe: pd.DataFrame) -> None:
```

```
    bandwidth = estimate_bandwidth(dataframe, quantile=0.1, n_samples=100)
```

```
    clusters = MeanShift(bandwidth=bandwidth, bin_seeding=True).fit(dataframe)
```

```
    sns.scatterplot(data=dataframe, x="low", y="high", hue=clusters.labels_)
```

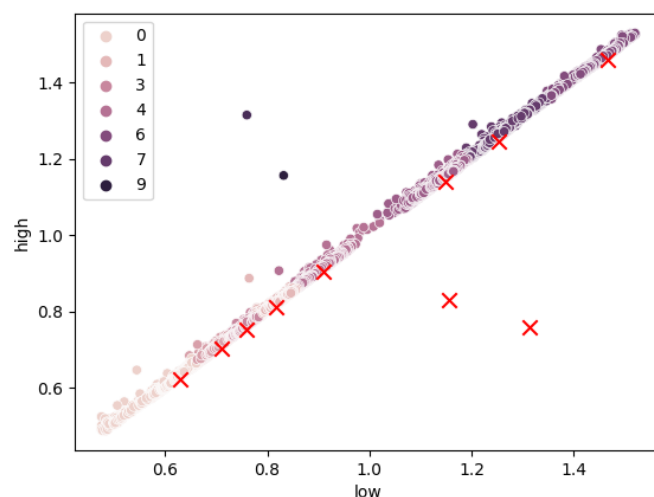
```
    plt.scatter(
```

```
        clusters.cluster_centers_[ :, 0], clusters.cluster_centers_[ :, 1], marker='x', c="r",  
        s=80, label="centroids"
```

```
)
```

```
    plt.savefig('mean_shift_cluster.png')
```

```
    plt.show()
```



Σχήμα 3: Αποτελέσματα MeanShift

**1.3.3. Σχολιασμός αποτελεσμάτων.** Τα αποτελέσματα των δύο αλγορίθμων παρουσιάζονται στα Σχήματα 2 και 3. Παρότι οι δύο αλγόριθμοι είναι αρκετά διαφορετικοί μεταξύ τους, τα αποτελέσματα είναι σχεδόν ίδια. Ο KMeans έχει χωρίσει τα clusters σε 4, ενώ ο MeanShift σε 10. Σε ένα πιο πρακτικό σενάριο, μοιάζει ο KMeans να είναι πιο ελεγχόμενος, ενώ ο MeanShift πιο πιστός στα πραγματικά δεδομένα. Για παράδειγμα, στον τομέα της Επεξεργασίας Εικόνας, ο KMeans θεωρείτε καλύτερος εφόσον παράγει παρόμοια αποτελέσματα αρκετά πιο γρήγορα από τον MeanShift.

## 2. Θεωρητικό υπόβαθρο της διαχείρισης γνώσης

### 2.1. Διαχωρισμός δεδομένων, πληροφορίας και γνώσης.

“Δεδομένα: ουσιαστικοποιημένο ουδέτερο πληθυντικού της μετοχής δεδομένος, έχει δύο σημασίες:

1. στοιχεία, πληροφορίες, σε δυαδική μορφή που εισάγονται προς επεξεργασία σε έναν ηλεκτρονικό υπολογιστή ή προβάλλονται ως έξοδος σε μια περιφερειακή συσκευή
2. στοιχεία, πληροφορίες, που έχουν λάβει δυαδική μορφή και έχουν αποθηκευτεί σε σκληρό δίσκο ή άλλο μέσο.

Στην πληροφορική ότι αφορά το λογισμικό (όχι το υλισμικό) είναι δεδομένα. Ακόμη και τα προγράμματα που εκτελούνται εμπεριέχονται σε αρχεία, τα εκτελέσιμα αρχεία, που τα δεδομένα τους είναι εντολές για το τι πρέπει να κάνει το πρόγραμμα. Όταν ξεκινάει ο υπολογιστής φορτώνει από προκαθορισμένη θέση του σκληρού δίσκου τα δεδομένα που του λένε που θα βρει αποθηκευμένα τα εκτελέσιμα αρχεία του λειτουργικού συστήματος. Μετά το υλισμικό ότι υπάρχει είναι δεδομένα.”

**Ορισμός «Δεδομένα» στο Βικιλεξικό**

Τα δεδομένα τα οποία αρχικά είχαν συλλεχθεί ήταν τα exchange rates όλων των νομισμάτων σε άλλα νομίσματα.

“Πληροφορία: θηλυκό ουσιαστικό που στην θεωρία της πληροφορίας είναι η τυχαία τιμή ή ο συνδυασμός τιμών εντός της επιτρεπόμενης πληροφοριακής εντροπίας χωρίς αναγκαστικά να μεταφέρει σημασιολογικό κώδικα.”

**Ορισμός «Πληροφορία» στο Βικιλεξικό**

Πληροφορία είναι μία συλλογή δεδομένων που έχει επεξεργαστεί για να μπορεί να γίνει περισσότερη χρήση του και να βγει ένα αποτέλεσμα. Στο συγκεκριμένο παράδειγμα, κρατήθηκαν μόνο τα δεδομένα του Ευρώ και έτσι το νέο αρχείο ήταν συγκεντρωμένο με πιο καθαρές πληροφορίες σχετικές με το θέμα.

“Γνώση: θηλυκό ουσιαστικό που είναι οι πληροφορίες που αποκτά κάποιος και οι παραστάσεις που σχηματίζει για τον κόσμο και τα πράγματα μετά από την νοητική επεξεργασία των εμπειρικών δεδομένων”

**Ορισμός «Γνώση» στο Βικιλεξικό**

Οι γνώσεις που μπορούν να θεωρηθούν από τα αποτελέσματα της εφαρμογής που μελετήθηκαν στο Κεφάλαιο 1.3.3. Με αυτά, μπορεί ο οποιοσδήποτε να παρατηρήσει ότι η διαφορά μεταξύ υψηλότερης και κατώτερης τιμής στα exchange offices παραμένουν παρόμοια κατά μέσο όρο στην πάροδο του χρόνου και μόνο σε συγκεκριμένες περιπτώσεις το θέμα γίνεται τραγικά κακό.

### 2.2. Κύκλος ζωής διαχείρισης γνώσης.

### 2.3. Μοντέλο αλλαγών του Kotter.