

Επιστήμη των δεδομένων

Γεώργιος Αλεξάκης 58093

14 Μαρτίου 2023

Περιεχόμενα

1	Εισαγωγή	4
2	Ταξινόμηση	6
3	Παλινδρόμηση	7
4	Ομαδοποίηση	8
5	Μείωση διαστάσεων	9
6	Ενισχυτική μάθηση	10
7	Ανίχνευση ανωμαλίας	11
8	Συμπέρασμα	12

Κατάλογος Σχημάτων

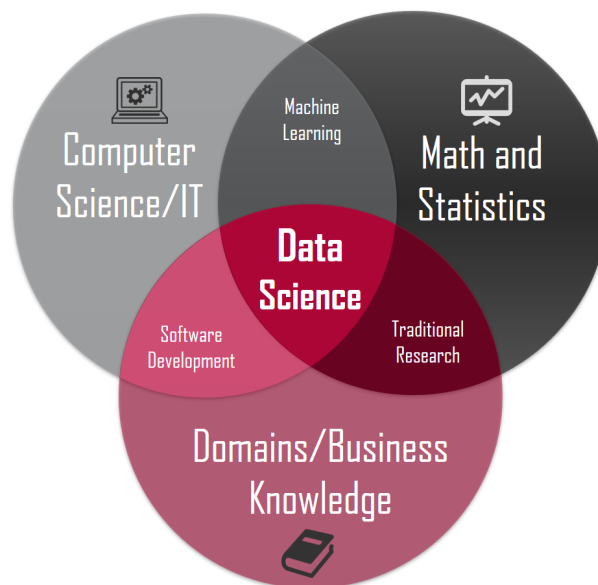
1	Γνώσεις που συνδυάζει η επιστήμη των δεδομένων	4
2	Αλγόριθμοι μηχανικής μάθησης	5

1 Εισαγωγή

Η επιστήμη των δεδομένων (Data science) είναι μια επιστήμη που αφορά την εξαγωγή γνώσης από δεδομένα (δομημένα ή αδόμητα) και για να το πετύχει αυτό συνδυάζει γνώσεις από διάφορες άλλες επιστήμες όπως:

- Μαθηματικά
- Στατιστική
- Προγραμματισμός
- Προγνωστική ανάλυση (Predictive analysis)
- Εξόρυξη δεδομένων (Data mining)
- Τεχνητή Νοημοσύνη (Artificial intelligence)
- Μηχανική μάθηση (Machine learning)

Επιπλέον, η επιστήμη των δεδομένων συνδυάζει τα παραπάνω με τη γνώση κάποιου ειδικού τομέα (όπως η ιατρική) με σκοπό να δώσει λύση σε ένα συγκεκριμένο πρόβλημα [1]



Σχήμα 1: Γνώσεις που συνδυάζει η επιστήμη των δεδομένων

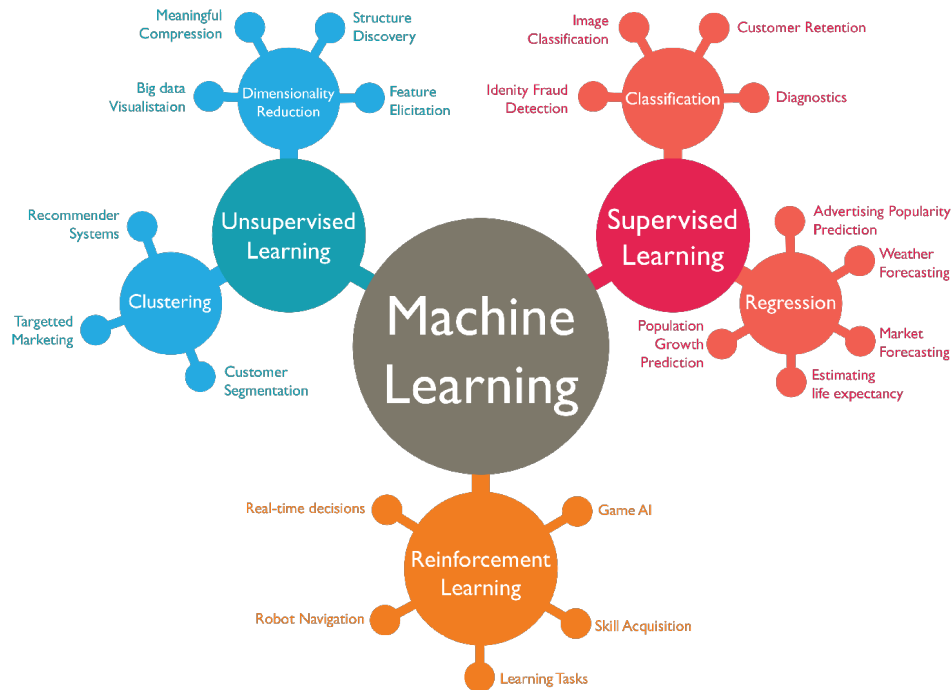
Σε αυτή την εργασία θα ασχοληθούμε κυρίως με την τεχνητή νοημοσύνη και τη μηχανική μάθηση και πιο συγκεκριμένα με τους αλγόριθμους που εφαρμόζονται στην επιστήμη των δεδομένων. Θα δοθεί αναλυτική εξήγηση για πολλούς χρήσιμους αλγόριθμους και θα υλοποιηθούν οι πιο σημαντικοί από αυτούς με τη χρήση της γλώσσας προγραμματισμού R η οποία είναι εξειδικευμένη για την επιστήμη των δεδομένων.

Οι αλγόριθμοι που θα εξετάσουμε μπορούν να χωριστούν σε πέντε ομάδες με βάση τη χρήση τους και αυτές είναι:

- Αλγόριθμοι ταξινόμησης (Classification)
- Αλγόριθμοι παλινδρόμησης (Regression)
- Αλγόριθμοι ομαδοποίησης (Clustering)

- Αλγόριθμοι μείωσης διαστάσεων (Dimensionality reduction)
- Αλγόριθμοι ενισχυτικής μάθησης (Reinforcement learning)
- Αλγόριθμοι ανίχνευσης ανωμαλίας (Anomaly detection)

Μας ενδιαφέρουν ιδιαίτερα οι πρώτες δύο ομάδες καθώς είναι αυτές που εμφανίζονται πιο συχνά. Πρόκειται για πολύ παρόμοιους αλγόριθμους οι οποίοι ανήκουν στους αλγόριθμους εποπτευόμενης μάθησης (Supervised learning) και χρησιμοποιούνται για να κάνουν προβλέψεις σύμφωνα με τα δεδομένα που τους έχουμε δώσει. Η διαφορά τους είναι ότι η ταξινόμηση προσπαθεί να προβλέψει την κλάση των καινούριων δεδομένων και να τα κατηγοριοποιήσει σύμφωνα με τις υπάρχουσες κατηγορίες, ενώ η παλινδρόμηση προσπαθεί να προβλέψει την τιμή κάποιου στοιχείου για τα καινούρια δεδομένα σύμφωνα με μια συνάρτηση που έφτιαξε από τα υπάρχοντα δεδομένα.



Σχήμα 2: Αλγόριθμοι μηχανικής μάθησης

2 Ταξινόμηση

Σε αυτή την ενότητα θα αναλύσουμε τους αλγορίθμους ταξινόμησης και θα δούμε τη χρήση τους.
Οι

3 Παλινδρόμηση

4 Ομαδοποίηση

5 Μείωση διαστάσεων

6 Ενισχυτική μάθηση

7 Ανίχνευση ανωμαλίας

8 Συμπέρασμα

References

- [1] Wikipedia. https://en.wikipedia.org/wiki/Data_science. Accessed: 2023-03-13.