

ΣΥΓΧΡΟΝΕΣ ΜΕΘΟΔΟΙ ΣΤΑΤΙΣΤΙΚΗΣ ΑΝΑΛΥΣΗΣ ΚΑΙ ΠΑΡΟΥΣΙΑΣΗΣ ΙΑΤΡΟ-ΒΙΟΛΟΓΙΚΩΝ ΚΑΙ ΕΠΙΔΗΜΙΟΛΟΓΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΧΡΗΣΗ R, PYTHON ΚΑΙ POWER BI

1. ΣΕ ΠΟΙΟΥΣ ΑΠΕΥΘΥΝΕΤΑΙ ΤΟ ΠΡΟΓΡΑΜΜΑ

Το Πρόγραμμα απευθύνεται σε απόφοιτους και φοιτητές Πανεπιστημιακών Ιδρυμάτων της ημεδαπής και της αλλοδαπής, ιδιώτες ή στελέχη επιχειρήσεων που στην εργασία τους χρειάζεται να αναλύουν δεδομένα και να χρησιμοποιούν R, Python ή/και Power BI, ερευνητές ή υποψήφιους διδάκτορες από διάφορα άλλα αντικείμενα που χρειάζονται για την έρευνά τους την χρήση των παραπάνω εργαλείων, στατιστικούς, μαθηματικούς, ιατρούς, κοινωνιολόγους, ψυχολόγους, διαιτολόγους/διατροφολόγους, επαγγελματίες στον χώρο της Πληροφορικής και του Marketing, Επιστήμονες Υγείας και όσους εργάζονται ή θέλουν να εργαστούν στον χώρο της Φαρμακοβιομηχανίας. Τέλος, οι μέθοδοι στατιστικής ανάλυσης και μηχανικής μάθησης που θα συζητηθούν είναι απαραίτητα προσόντα για κάθε ερευνητή, ανεξαρτήτως πεδίου, που ανήκει σε κάποια ερευνητική ομάδα, ώστε να μπορεί να κατανοεί και ερμηνεύει σωστά τα αποτελέσματα των στατιστικών διεργασιών καθώς και να διαβάζει με άνεση σχετικά άρθρα από την ελληνική και διεθνή Βιβλιογραφία και για επιχειρήσεις που επιθυμούν να εκμεταλλευτούν τη δύναμη των δεδομένων και να καταλήξουν σε data driven αποφάσεις.

2. ΣΚΟΠΟΣ ΠΡΟΓΡΑΜΜΑΤΟΣ

Σκοπός του προγράμματος αρχικά είναι η εισαγωγή στις βασικές έννοιες προγραμματισμού στην γλώσσα R και Python, ενώ στην συνέχεια η εκμάθησή τους για χρήση σε θέματα προχωρημένης περιγραφικής και επαγωγικής στατιστικής. Το συγκεκριμένο πρόγραμμα ξεκινάει από το μηδέν τις αντίστοιχες μεθοδολογίες, επομένως θα καλύψει πλήρως τον αρχάριο χρήστη από οποιονδήποτε ερευνητικό τομέα και αν προέρχεται, αλλά θα προσφέρει και στον έμπειρο τις κατάλληλες κατευθύνσεις, μέσα από την μελέτη περιπτώσεων – πραγματικά ιατροβιολογικά προβλήματα - και την ενδεικτική βιβλιογραφία, ώστε να επεκτείνει τις γνώσεις του.

Στην συνέχεια, θα συζητηθούν προχωρημένα θέματα στατιστικής ανάλυσης (Μηχανική Μάθηση - ML) ενώ θα τρέξουν εφαρμογές σε πραγματικά δεδομένα. Τέλος, ο εκπαιδευόμενος θα μπορεί να οπτικοποιήσει και να παρακολουθεί μεγάλες βάσεις δεδομένων με τρόπο διαδραστικό όχι μόνο για τον ίδιο ως χειρίστη αλλά και για κάθε μεμονωμένο χρήστη στον οποίο απευθύνεται η πληροφορία, μέσω του πιο διαδεδομένου εργαλείου διαχείρισης και οπτικοποίησης δεδομένων, το Power BI.

3. ΠΡΟΣΔΟΚΩΜΕΝΑ ΜΑΘΗΣΙΑΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Οι εκπαιδευόμενοι θα μάθουν να προγραμματίζουν και να χρησιμοποιούν τις γλώσσες R και Python επαρκώς για την ανάλυση πάσης φύσεως δεδομένων. Οι εφαρμογές παρόλο που θα γίνουν σε Ιατροβιολογικά και Επιδημιολογικά δεδομένα, η γνώση που θα λάβουν οι εκπαιδευόμενοι μπορεί να εφαρμοστεί αντίστοιχα και σε άλλα ερευνητικά πεδία. Ειδικότερα, σε όλους τους Εκπαιδευόμενους με κατεύθυνση τη στατιστική ή τους επιστήμονες που ανήκουν σε μια ερευνητική ομάδα, θα δοθεί μια εξειδίκευση στο χώρο της υγείας, ενώ θα αναπτύξουν δεξιότητες ανάλυσης και επεξεργασίας πάσης φύσεως δεδομένων. Τέλος, θα μάθουν να διαχειρίζονται και να οπτικοποιούν μεγάλες βάσεις δεδομένων και να δημιουργούν αυτοματοποιημένες αναλύσεις. Η απόκτηση των δεξιοτήτων θα στοχεύσει στον χώρο της υγείας και τα δεδομένα που παράγονται γύρω και μέσα από αυτόν, αλλά η γνώση που θα λάβουν οι εκπαιδευόμενοι μέσα από την εκμάθηση του Power BI μπορεί να εφαρμοστεί σε οποιονδήποτε επιστημονικό τομέα (π.χ Business Intelligence, Marketing κτλ).

4. ΠΕΡΙΕΧΟΜΕΝΑ ΠΡΟΓΡΑΜΜΑΤΟΣ

Το αναλυτικό περιεχόμενο του προγράμματος έχει δομηθεί κατά τρόπο που να καλύπτονται οι γνώσεις και οι δεξιότητες που θεωρείται αναγκαίο να διαθέτει ο εκπαιδευόμενος με βάση το Επαγγελματικό Περίγραμμά του, καθώς επίσης και τα προβλεπόμενα μαθησιακά αποτελέσματα του προγράμματος σπουδών. Στη συνέχεια παρατίθενται συνοπτικά οι ενότητες του προγράμματος.

**Οι μελέτες περίπτωσης είναι ενδεικτικές του συνόλου των προβλημάτων που θα μελετηθούν τελικά στο σεμινάριο. Τα προβλήματα που θα εξεταστούν θα εξαρτηθούν από τις ανάγκες και το επαγγελματικό υπόβαθρο των εκπαιδευομένων.*

Περιγραφή Θεματικής Ενότητας Part_A	Ώρες
<p>Ενότητα 1: Στατιστική: Η Επιστήμη των Δεδομένων</p> <p>Περιγραφή: Στην 1η Ενότητα θα κάνουμε μια Ιστορική Αναδρομή στην Στατιστική, ώστε ο εκπαιδευόμενος να κατανοήσει σε βάθος τις ρίζες της Επιστήμης.</p>	1
<p>Ενότητα 2: Εισαγωγή στη Στατιστική Γλώσσα Προγραμματισμού R</p> <p>ΥποΕνότητες:</p> <ol style="list-style-type: none"> 1.1 Ιστορική αναδρομή της στατιστικής γλώσσας R 1.2 Οδηγός εγκατάστασης σε Windows και Mac 1.3 Το περιβάλλον του R - Studio 1.4 Τύποι και Δομές Δεδομένων 1.5 Αποθήκευση και ανάκτηση δεδομένων 1.6 Δομές Ελέγχου και Επανάληψης 1.7 Η οικογένεια εντολών Apply 1.8 Διαχείριση δεδομένων με τα πακέτα dplyr & tidyr 1.9 Βασικά Διαγράμματα στην R 1.10 Οπτικοποίηση δεδομένων με το πακέτο ggplot2 	6
<p>Ενότητα 3: Εισαγωγή στη Στατιστική Γλώσσα Προγραμματισμού Python</p> <p>ΥποΕνότητες:</p> <ol style="list-style-type: none"> 1.1 Ιστορική αναδρομή της γλώσσας προγραμματισμού Python 1.2 Πλεονεκτήματα και βασικά χαρακτηριστικά 1.3 Οδηγός εγκατάστασης σε Windows και Mac 1.4 Το περιβάλλον του Jupyter Notebook 1.5 Αρθρώματα και Βιβλιοθήκες στην Python 1.6 Τύποι και δομές δεδομένων 1.7 Δομές ελέγχου και επανάληψης 1.8 Αποθήκευση και ανάκτηση δεδομένων 	6

<p>1.9 Η βιβλιοθήκη ανάλυσης δεδομένων Pandas 1.10 Η μαθηματική βιβλιοθήκη NumPy 1.11 Οπτικοποίηση δεδομένων με τις βιβλιοθήκες Matplotlib και seaborn</p>	
<p>Ενότητα 4: Μέθοδοι δειγματοληψίας και καθορισμός κατάλληλου μεγέθους δείγματος</p> <p>Περιγραφή: Στην 4η Διδακτική Ενότητα ο εκπαιδευόμενος θα γνωρίσει όλες τις διαθέσιμες μεθόδους Δειγματοληψίας και πως να επιλέγει την καταλληλότερη μέθοδο για την δική του μελέτη. Επιπλέον θα εκπαιδευτεί στην R και την Python στην διαδικασία επιλογής κατάλληλου μεγέθους δείγματος.</p>	2
<p>Ενότητα 5: Περιγραφική Στατιστική ανάλυση και παρουσίαση ποιοτικών δεδομένων</p> <p>Περιγραφή: Στόχος της 5ης Διδακτικής Ενότητας είναι να μάθει τον εκπαιδευόμενο να εξάγει από ποιοτικά δεδομένα την πληροφορία που χρειάζεται με γραφικές και αριθμητικές τεχνικές στην R και την Python.</p>	3
<p>Ενότητα 6: Περιγραφική Στατιστική ανάλυση και παρουσίαση ποσοτικών δεδομένων</p> <p>Περιγραφή: Στόχος της 6ης Διδακτικής Ενότητας είναι να μάθει τον εκπαιδευόμενο να εξάγει από ποσοτικά δεδομένα την πληροφορία που χρειάζεται με γραφικές και αριθμητικές τεχνικές στην R και την Python.</p>	3
<p>Ενότητα 7: Στατιστική ανάλυση και οπτικοποίηση πολυμεταβλητών δεδομένων</p> <p>Περιγραφή:</p>	2

<p>Στόχος της 7ης Διδακτικής Ενότητας είναι να μάθει τον εκπαιδευόμενο να εξάγει χρήσιμη πληροφορία χρησιμοποιώντας ταυτόχρονα περισσότερες από 1 μεταβλητές.</p>	
<p>Ενότητα 8: Κατανομές Πιθανοτήτων</p> <p>Περιγραφή: Όταν μετράμε για παράδειγμα κάποιους βιοχημικούς δείκτες καρδιαγγειακών νοσημάτων (LDL -χοληστερίνη, τροπονίνες, νατριοδιουρητικά πεπτίδια) ή κάποια κοινωνικο-δημογραφικά χαρακτηριστικά (ηλικία, αριθμός παιδιών, συνολικά έτη εκπαίδευσης, εισόδημα) το αποτέλεσμα που προκύπτει το αναπαριστούμε με μία αριθμητική τιμή της αντίστοιχης μεταβλητής. Όταν οι τιμές αυτές είναι αποτέλεσμα ενός πειράματος τύχης, η μεταβλητή αυτή καλείται τυχαία μεταβλητή. Οι τυχαίες μεταβλητές διακρίνονται σε διακριτές και συνεχείς. Στόχος της 8ης Διδακτικής Ενότητας είναι να εξοικειώσει τον εκπαιδευόμενο με τις βασικές διακριτές και συνεχείς κατανομές πιθανοτήτων, μέσα από την επίλυση σύγχρονων προβλημάτων στην R και την Python.</p>	2
<p>Ενότητα 9: Εκτιμητική και Έλεγχος Στατιστικών Υποθέσεων: Βασικές Παραμετρικές και μη Παραμετρικές Μέθοδοι</p> <p>Περιγραφή: Η επαγωγική στατιστική (inferential statistics) περιλαμβάνει τεχνικές που επιτρέπουν την ανάλυση δεδομένων, ώστε να γίνει εφικτή η εξαγωγή χρήσιμων και γενικεύσιμων συμπερασμάτων για τους πληθυσμούς, με βάση τις πληροφορίες που συλλέγονται από τα δείγματα. Στόχος της 9ης Διδακτικής Ενότητας είναι να εισάγει τον εκπαιδευόμενο στους δύο υποκλάδους της Επαγωγικής Στατιστικής: την Εκτιμητική και τον Έλεγχο Στατιστικών Υποθέσεων. Θα γίνει εφαρμογή σε πραγματικά δεδομένα με χρήση R και Python</p>	6

<p>Μελέτη περίπτωσης*: <u>Ανάλυση Επιπέδων Γονιδιακής Έκφρασης:</u> Έστω ένα σύνολο δεδομένων γονιδιακής έκφρασης που προέρχεται από μια μελέτη φαρμακολογικής παρέμβασης σε κύτταρα ποντικών που υποφέρουν από φλεγμονώδη αρθρίτιδα. Πόσα και ποια γονίδια για τα οποία η ασθένεια οδηγεί σε στατιστικά σημαντική μεταβολή της έκφρασης? Ποια θεραπεία είναι πιο αποτελεσματική σε επίπεδο γονιδιακής έκφρασης?</p>	
<p>Ενότητα 10: Ανάλυση κατηγορικών δεδομένων</p> <p>Περιγραφή: Στόχος της 10ης Ενότητας είναι η ανάλυση κατηγορικών δεδομένων με τον έλεγχο ανεξαρτησίας Chi – Square του Pearson, τον ακριβή έλεγχο το Fisher, και το McNemar τεστ. Θα γίνει εφαρμογή σε πραγματικά δεδομένα με χρήση R και Python.</p> <p>Μελέτη περίπτωσης*: Θα μελετηθεί η σχέση μεταξύ του Φύλου και της Κατάστασης Υγείας των ασθενών σε μια Κλινική Μελέτη.</p>	2
<p>Ενότητα 11: Είδη Επιδημιολογικών μελετών και αξιολόγηση διαγνωστικής ικανότητας</p> <p>Περιγραφή: Στην 11^η Ενότητα θα συζητηθούν τα βασικότερα είδη των Επιδημιολογικών Μελετών, όπως μελέτες επιπολασμού, μελέτες κοορτής, μελέτες ασθενών - μαρτύρων και παρεμβατικές μελέτες (π.χ. κλινικές δοκιμές), καθώς και βασικά μέτρα συσχέτισης, όπως ο σχετικός κίνδυνος και ο σχετικός λόγος, ο συντελεστής Φ και C του Pearson και ο συντελεστής V του Cramer. Στόχος αυτής της υπο-ενότητας είναι η εξοικείωση του εκπαιδευόμενου με τον επιστημονικό χώρο της <i>Επιδημιολογίας</i> και της <i>Δημόσιας Υγείας</i>.</p>	2

<p>Επιπλέον θα γίνει μια σύντομη συζήτηση πάνω σε θέματα αξιολόγησης διαγνωστικής ικανότητας (Ευαισθησία – Ειδικότητα, Καμπύλη ROC κτλ.).</p> <p>Μελέτη περίπτωσης*: Θα εξεταστεί αν η Γαλακτική αφυδρογονάση (LDH) είναι καλός δείκτης διάγνωσης του μυοκαρδίου έναντι ασταθούς στηθάγχης.</p>	
<p>Ενότητα 12: Δημιουργία, Διαχείριση, Επεξεργασία και Οπτικοποίηση βάσεων δεδομένων με χρήση του λογισμικού Power BI: Εφαρμογή στα δεδομένα της πανδημίας</p> <p>Περιγραφή: Στόχος της 12ης Διδακτικής Ενότητας είναι να μάθει τον εκπαιδευόμενο τεχνικές διαχείρισης και οπτικοποίησης μεγάλων όγκων δεδομένων μέσω εργαλείων Business Intelligence (Power BI). Ειδικότερα θα του μάθει εισαγωγή δεδομένων από διάφορες πηγές σε μοντέλα δεδομένων</p> <ol style="list-style-type: none"> 1. Δημιουργία σύνθετων υπολογισμών 2. Ανανέωση αποτελεσμάτων με αυτοματοποιημένες διαδικασίες 3. Παρουσίαση αναλύσεων για χρήση σε κινητές συσκευές η Tablet <p>Μελέτη περίπτωσης*: Θα πραγματοποιηθεί μια διαδραστική απεικόνιση των δεδομένων της πανδημίας με χρήση του Power BI ενώ θα παρουσιάσουμε βήμα – βήμα την διαδικασία επεξεργασίας. Στόχος είναι να δημιουργηθεί ένα dashboard με τρόπο κατανοητό ώστε ο εκπαιδευόμενος μετά το πέρας αυτής της διάλεξης να μπορεί να παρουσιάσει με ευκολία και τα δικά του δεδομένα.</p>	4

<p>Ενότητα 13: Μονοπαραγοντική Ανάλυση Διακύμανσης και Συνδιακύμανσης</p> <p>Περιγραφή: Η Ανάλυση Διασποράς αποτελεί επέκταση των στατιστικών ελέγχων υποθέσεων t και z που μελετήσαμε στην 9η Ενότητα. Η Ανάλυση Διακύμανσης (ANOVA) και Συνδιακύμανσης (ANCOVA) χρησιμοποιείται έντονα στις Επιστήμες Υγείας για την μελέτη των σχέσεων μεταξύ των μεταβλητών επομένως στην 13η Διδακτική Ενότητα θα δούμε κάποια τέτοια παραδείγματα. Θα γίνει εφαρμογή σε πραγματικά δεδομένα με χρήση R και Python.</p> <p>Μελέτη περίπτωση*: Μας ενδιαφέρει να ελέγξουμε εάν διαφέρει ο αριθμός των αιμοπεταλίων μεταξύ ατόμων διαφόρων τύπων μάζας σώματος. Αλλάζει η συγκεκριμένη σχέση μεταξύ ανδρών και γυναικών?</p>	4
<p>Ενότητα 14: Ανάλυση Διακύμανσης για Επαναλαμβανόμενες μετρήσεις</p> <p>Περιγραφή: Μια από τις προϋποθέσεις που πρέπει να πληρούνται για να είναι αξιόπιστες οι Μέθοδοι Ανάλυσης Διασποράς που εξετάστηκαν στην 13η Ενότητα είναι ότι οι συγκρινόμενες ομάδες θα πρέπει να είναι ανεξάρτητες μεταξύ τους. Στην Ιατρική Επιστήμη είναι συχνό φαινόμενο οι υπό μελέτη ομάδες να μην ικανοποιούν αυτό το κριτήριο.</p> <p>Μελέτη περίπτωση*: Μας ενδιαφέρει να ελέγξουμε την επίδοση ενός ασθενή κάτω από 3 διαφορετικές θεραπευτικές αγωγές</p>	2

<p>Ενότητα 15: Ανάλυση συσχέτισης και μερικής συσχέτισης</p> <p>Περιγραφή: Στόχος της 15ης Διδακτικής Ενότητας είναι η μελέτη της σχέσης μεταξύ 2 ποσοτικών μεταβλητών. Για την εκτίμηση και τον έλεγχο της συσχέτισης θα ασχοληθούμε με την γραμμική συσχέτιση του Pearson, με τον συντελεστή συσχέτισης rho του Spearman και με τον συντελεστή tau του Kendall. Θα γίνει εφαρμογή σε πραγματικά δεδομένα με χρήση R και Python</p> <p>Μελέτη περίπτωση*: Να ελεγχθεί η συσχέτιση μεταξύ του Δείκτη Μάζας Σώματος και της Ηλικίας. Με ποιον τρόπο επηρεάζουν την παραπάνω σχέση ένας ή περισσότεροι βιοχημικές παράμετροι (σάκχαρο, ινσουλίνη, ουρία, κρεατινίνη κλπ)?</p>	2
<p>Ενότητα 16: Εισαγωγή στην Μηχανική Μάθηση</p> <p>Περιγραφή: Η Μηχανική Μάθηση (Machine Learning) είναι μια περιοχή της Τεχνητής Νοημοσύνης (AI) και υποπεδίο της Επιστήμης των Υπολογιστών που αναπτύχθηκε από την ανάγκη ανάλυσης αναγνώρισης προτύπων (Pattern Recognition Analysis). Στην 16η Ενότητα θα γίνει αναφορά στις μεθόδους της Μηχανικής Μάθησης ενώ θα αναφερθούμε στο ποια είναι η χρήση αυτών σε σύγχρονα προβλήματα στον χώρο του Φαρμάκου και της Υγείας.</p> <p>Συγκεκριμένα, θα αναφερθούμε στις <i>Επιβλεπόμενες και μη Επιβλεπόμενες Μεθόδους Ανάλυσης</i> (Παλινδρόμηση - Ταξινόμηση - Ομαδοποίηση), στις <i>Μεθόδους Μείωσης της Διαστασιμότητας</i>, στις <i>Μετρικές Αξιολόγησης</i> των παραπάνω μεθόδων και στην <i>Ενισχυτική Μάθηση</i>.</p> <p>Ενδεικτικές Εφαρμογές: Drug Discovery Clinical Trials Personalized Medicine Medical Diagnosis</p>	2

<p>1^ο Παράδειγμα: Αλγόριθμοι Μηχανικής Μάθησης στα νοσοκομεία αυξάνουν την εξοικονόμηση κόστους, βελτιώνουν την ικανοποίηση των ασθενών και ικανοποιούν τις ανάγκες του προσωπικού τους.</p> <p>2^ο Παράδειγμα: Υποθέτοντας ότι έχουμε ένα σύνολο δεδομένων από εικόνες από ακτινογραφίες ασθενών, θα χρησιμοποιήσουμε τα παραπάνω μοντέλα για να καταλήξουμε αν ο εξεταζόμενος πάσχει από μια ασθένεια ή όχι</p>	
<p>Ενότητα 17: Ανάλυση Παλινδρόμησης</p> <p>Περιγραφή: Η παλινδρόμηση είναι μια τεχνική μοντελοποίησης για την έρευνα της συσχέτισης μεταξύ μιας εξαρτημένης μεταβλητής, και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Το αποτέλεσμα του μοντέλου θα είναι μια πρόβλεψη της εξαρτημένης μεταβλητής που οι τιμές της ανήκουν στους συνεχής αριθμούς.</p> <p>Μελέτη περίπτωση*: Να κατασκευαστεί ένα μοντέλο πρόβλεψης των επιπέδων ολικής χοληστερόλης από την ηλικία, το Φύλο, τον Δείκτη Μάζας Σώματος, τις Καπνιστικές Συνήθειες και το ημερήσιο επίπεδο κατανάλωσης αλκοόλ</p>	4
<p>Ενότητα 18: Ιεραρχική Παλινδρόμηση και επιλογή βέλτιστου μοντέλου</p> <p>Περιγραφή: Το πρόβλημα εύρεσης του καλύτερου μοντέλου συνίσταται στην επιλογή ενός κατάλληλου υποσυνόλου ανεξάρτητων μεταβλητών με σκοπό τη μέγιστη ερμηνευτική ικανότητα του μοντέλου. Στην 18η Διδακτική Ενότητα θα εξετάσουμε την Ιεραρχική Παλινδρόμηση ενώ θα μελετήσουμε επαναληπτικές τεχνικές επιλογής βέλτιστων υποσυνόλων.</p> <p>Μελέτη περίπτωση*: Μας ενδιαφέρει να επιλέξουμε το βέλτιστο μοντέλο για την πρόβλεψη των τιμών της συστολικής πίεσης</p>	1

χρησιμοποιώντας ένα σύνολο πιθανών προβλεπτικών παραγόντων.	
<p>Ενότητα 19: Ειδικά θέματα στην Παλινδρόμηση – Μη γραμμικά μοντέλα</p> <p>Περιγραφή: Στην 18η Ενότητα θα συζητηθούν περιληπτικά κάποια ειδικά θέματα στην Παλινδρόμηση όπως η Παλινδρόμηση Lasso και Ridge, η Πολυωνυμική Παλινδρόμηση, η Παλινδρόμηση Poisson, τα Γενικευμένα Αθροιστικά Μοντέλα (GAM) , η Ανάλυση Επιβίωσης (Μοντέλο Παλινδρόμησης Cox) κ.α.</p>	2
<p>Ενότητα 20: Λογιστική Παλινδρόμηση</p> <p>Περιγραφή: Στις Διδακτικές Ενότητες 17 και 18 ασχοληθήκαμε με την εκτίμηση της επίδρασης που ασκούν διάφοροι παράγοντες στην υπό μελέτη έκβαση. Η έκβαση αφορούσε ένα ποσοτικό χαρακτηριστικό, ενώ πολλές φορές στην Βιοϊατρική Έρευνα μελετώνται παράγοντες που επηρεάζουν ένα ποιοτικό χαρακτηριστικό. Η Λογιστική Παλινδρόμηση αποτελεί ένα μοντέλο ταξινόμησης των τιμών μιας ποιοτικής μεταβλητής απόκρισης με βάση ένα πλήθος παραγόντων (ποσοτικών ή ποιοτικών) που επιλέγονται για το σκοπό αυτό.</p> <p>Μελέτη περίπτωση*: <u>Διάγνωση Καρκίνου του Μαστού</u> Η έγκαιρη και σωστή διάγνωση μιας πάθησης είναι ένα από τα σημαντικότερα προβλήματα της Ιατρικής που έρχεται όμως να λύσει/προσεγγίσει η Επιστήμη των Δεδομένων.</p> <p>Έστω ένα σύνολο δεδομένων* με ασθενείς με καρκίνο του μαστού που περιέχει λεπτομερείς μετρήσεις για το σχήμα των πυρήνων κυττάρων του όγκου από βιοψίες παρακέντησης λεπτής βελόνης (FNA). Θα δημιουργήσουμε ένα μοντέλο πρόβλεψης των δύο βασικών τύπων καρκίνου (μεταστατικού - μη μεταστατικού).</p> <p>*Θα χρησιμοποιηθεί το δημοσιευμένο σύνολο δεδομένων που συλλέχθηκε και αναλύθηκε από το Πανεπιστήμιο Wisconsin.</p>	2

Ενότητα 21: Συγγραφή ενός ερευνητικού άρθρου	2
Περιγραφή: Οι εκπαιδευόμενοι θα συγγράψουν με την βοήθεια του εκπαιδευτή ένα άρθρο αναλύοντας πραγματικά Ιατρο - Βιολογικά Δεδομένα.	

Περιγραφή Θεματικής Ενότητας Part_B

Ενότητα 22: Μοντέλα Μηχανικής Μάθησης I: Naive Bayes

Περιγραφή: Στην ενότητα αυτή, θα συζητηθούν διάφοροι αλγόριθμοι μηχανικής μάθησης τόσο για μοντέλα παλινδρόμησης όσο και ταξινόμησης. Βασικός σκοπός αυτής της ενότητας είναι να μάθουν οι εκπαιδευόμενοι τις βασικές αρχές των αλγορίθμων (δέντρα αποφάσεων, Random Forests, SVMs, naive bayes), να δουν τις διαφορές στις επιδόσεις με αντίστοιχα στατιστικά μοντέλα όπως επίσης και να μπορούν να προγραμματίζουν αυτά τα μοντέλα με τις κατάλληλες βιβλιοθήκες με τη χρήση python.

Μελέτη περίπτωσης: Θα χρησιμοποιηθούν τα ίδια δεδομένα με τις ενότητες 17 και 20 για να μπορέσουν οι μαθητές να καταλάβουν τις διαφορές στις επιδόσεις των μοντέλων.

Ενότητα 23: Μοντέλα Μηχανικής Μάθησης II: Δέντρα Αποφάσεων

Περιγραφή: Στην ενότητα αυτή, θα παρουσιαστεί ο αλγόριθμος μηχανικής μάθησης των δέντρων αποφάσεων. Ο αλγόριθμος αυτός μπορεί να χρησιμοποιηθεί και για προβλήματα παλινδρόμησης όπως και ταξινόμησης. Οι εκπαιδευόμενοι θα διδαχθούν τις βασικές αρχές του αλγορίθμου, πως επιλέγει το τελικό αποτέλεσμα και τα κριτήρια διάσπασης των δέντρων. Τέλος θα γίνει εφαρμογή του αλγορίθμου σε python πάνω σε πραγματικά βιολογικά δεδομένα.

Μελέτη περίπτωσης: Θα χρησιμοποιηθούν τα ίδια δεδομένα με τις ενότητες 17 και 20 για να μπορέσουν οι μαθητές να καταλάβουν τις διαφορές στις επιδόσεις των μοντέλων.

Ενότητα 24: Μοντέλα Μηχανικής Μάθησης III: Random Forests

Περιγραφή: Σε συνέχεια της προηγούμενης ενότητας, τα Random Forests, είναι ένας αλγόριθμος που

αποτελείται από πολλά δέντρα αποφάσεων (ensemble technique). Χρησιμοποιείται για προβλήματα παλινδρόμησης και ταξινόμησης. Οι εκπαιδευόμενοι θα μάθουν πως από πολλά διαφορετικά δέντρα καταλήγουμε σε ένα αποτέλεσμα, τις μεταβλητές που περιέχει το συγκεκριμένο μοντέλο, όπως επίσης θα γίνει εφαρμογή πάνω σε πραγματικά βιολογικά δεδομένα με τη χρήση της rpythοn.

Μελέτη περίπτωσης: Θα χρησιμοποιηθούν τα ίδια δεδομένα με τις ενότητες 17 και 20 για να μπορέσουν οι μαθητές να καταλάβουν τις διαφορές στις επιδόσεις των μοντέλων.

Ενότητα 25: Μοντέλα Μηχανικής Μάθησης IV: Support vector machines (Μηχανές Διανυσματικής Υποστήριξης)

Περιγραφή: Στην ενότητα αυτή θα εξεταστεί ο αλγόριθμος των Support Vector Machines. Το μοντέλο αυτό χρησιμοποιείται τόσο για προβλήματα παλινδρόμησης όσο και ταξινόμησης. Οι εκπαιδευόμενοι θα διδαχθούν τα μαθηματικά πίσω από τον αλγόριθμο, πως καταλήγει στην τελική απόφαση, ποιές είναι οι βασικοί παράμετροι που πρέπει να οριστούν στο μοντέλο και πως το επηρεάζουν. Τέλος, θα γίνει εφαρμογή του μοντέλου σε πραγματικά βιολογικά δεδομένα με τη χρήση της rpythοn.

Μελέτη περίπτωσης: Θα χρησιμοποιηθούν τα ίδια δεδομένα με τις ενότητες 17 και 20 για να μπορέσουν οι μαθητές να καταλάβουν τις διαφορές στις επιδόσεις των μοντέλων.

Ενότητα 26: Αξιολόγηση απόδοσης μοντέλων

Περιγραφή: Στην ενότητα 26 θα μελετήσουμε πώς μπορούμε να αξιολογήσουμε τα μοντέλα μας με βάση τις προβλέψεις που έκαναν. Θα εξετάσουμε πως πρέπει να χωρίζουμε τα δεδομένα μας (train-test split) καθώς και τις μετρικές που χρησιμοποιούνται για την αξιολόγηση των μοντέλων

Μελέτη περίπτωσης: Θα χρησιμοποιηθούν τα ίδια δεδομένα με τις ενότητες 17 και 20 για να μπορέσουν οι μαθητές να καταλάβουν τις διαφορές στις επιδόσεις των μοντέλων

Ενότητα 27: Προσδιορισμός υπερπαραμέτρων

Περιγραφή: Οι αλγόριθμοι μηχανικής μάθησης έχουν κάποιες μεταβλητές στα μοντέλα τους (υπερπαραμέτροι) που αναλόγως την τιμή που θα ορίσουμε, επηρεάζεται η απόδοση των μοντέλων. Στην ενότητα αυτή θα αναλυθούν οι υπερπαραμέτροι κάθε μοντέλου καθώς και τους τρόπους με τους οποίους μπορούμε να τους ορίσουμε έτσι ώστε τα μοντέλα μας να έχουν την καλύτερη δυνατή απόδοση.

Μελέτη περίπτωσης: Θα χρησιμοποιηθούν τα ίδια δεδομένα με τις ενότητες 17 και 20 για να μπορέσουν οι μαθητές να καταλάβουν τις διαφορές στις επιδόσεις των μοντέλων

Ενότητα 28: Εισαγωγή στα νευρωνικά δίκτυα και τη βαθιά μάθηση (deep learning)

Περιγραφή: Το νευρωνικό δίκτυο είναι ένα δίκτυο από απλούς υπολογιστικούς κόμβους (νευρώνες, νευρώνια), διασυνδεδεμένους μεταξύ τους. Είναι εμπνευσμένο από το Κεντρικό Νευρικό Σύστημα (ΚΝΣ), το οποίο προσπαθεί να προσομοιώσει. Είναι η πιο απλή μορφή μοντέλων Τεχνητής Νοημοσύνης. Στην ενότητα αυτή οι εκπαιδευόμενοι θα διδαχθούν πως την πιο απλή αρχιτεκτονική νευρωνικών δικτύων, τι προβλήματα μπορούν να επιλύσουν και πώς λειτουργούν

Ενότητα 29: Εκπαίδευση νευρωνικών δικτύων

Περιγραφή: Στην ενότητα αυτή οι εκπαιδευόμενοι θα καταλάβουν σε βάθος πως εκπαιδεύεται ένα νευρωνικό δίκτυο και θα αναλυθούν μαθηματικά ο αλγόριθμος backpropagation με απλά παραδείγματα. Επίσης, θα υλοποιηθεί από την αρχή ο κώδικας για ένα απλό νευρωνικό δίκτυο και θα αναφερθούν οι πιο γνωστές βιβλιοθήκες σε rython που χρησιμοποιούνται στη βιομηχανία (tensorflow, pytorch, keras).

Μελέτη περίπτωσης: Θα χρησιμοποιηθούν τα ίδια δεδομένα με τις ενότητες 17 και 20 για να μπορέσουν οι μαθητές να καταλάβουν τις διαφορές στις επιδόσεις των μοντέλων

Ενότητα 30: Εισαγωγή στα Convolutional Neural Networks

Περιγραφή: Τα convolutional neural networks, είναι μοντέλα της βαθιάς μάθησης, τα οποία χρησιμοποιούνται για να κάνουμε προβλέψεις σε δεδομένα που αποτελούνται από εικόνες. Στην ενότητα αυτή οι εκπαιδευόμενοι θα διδαχθούν τις βασικές μαθηματικές πράξεις που πραγματοποιούνται πίσω από τα μοντέλα καθώς και ποιες βιβλιοθήκες χρησιμοποιούνται για την εύκολη υλοποίησή τους.

Μελέτη περίπτωσης: Υποθέτοντας ότι έχουμε ένα σύνολο δεδομένων από εικόνες από ακτινογραφίες ασθενών, θα χρησιμοποιήσουμε τα παραπάνω μοντέλα για να καταλήξουμε αν ο εξεταζόμενος πάσχει από μια ασθένεια ή όχι

Ενότητα 31: Εισαγωγή στον τομέα του Natural Language Preprocessing (NLP)

Περιγραφή: Στην ενότητα αυτή οι εκπαιδευόμενοι θα διδαχθούν βασικές τεχνικές του NLP. Είναι γνωστό ότι μεγάλο μέρος της επιστημονικής γνώσης βρίσκεται γραμμένο σε άρθρα και σε μορφή που οι υπολογιστές δεν μπορούν να καταλάβουν. Ιδιαίτερο ενδιαφέρον προκαλεί ο τρόπος που μπορούμε να μετατρέψουμε τις λέξεις σε μορφή κατανοητή για τους υπολογιστές έτσι ώστε να χρησιμοποιήσουμε μοντέλα τεχνητής νοημοσύνης για να κάνουμε προβλέψεις. Ένα κλασικό παράδειγμα είναι η εύρεση λέξεων που μοιάζουν μεταξύ τους με βάση την απόσταση της αριθμητικής αναπαράστασης τους.

Ενότητα 32: Εισαγωγή στους γράφους

Περιγραφή: Στην ενότητα αυτή οι εκπαιδευόμενοι θα εισαχθούν στην έννοια των γράφων και πως μπορούμε να αναπαραστήσουμε δεδομένα με τη συγκεκριμένη δομή. Επίσης θα αναφερθούν βασικά χαρακτηριστικά και είδη των γράφων όπως και κάποιες ιδιότητες των ακμών και των κορυφών που απαρτίζουν έναν γράφο.

Ενότητα 33: Χρήση μοντέλων μηχανικής μάθησης για ανακάλυψη φαρμάκων

Part_C Επιδημιολογία και Δημόσια Υγεία με τη χρήση σύγχρονων εργαλείων ανάλυσης και της γλώσσας προγραμματισμού R: Εφαρμογές σε Ιατροβιολογικά δεδομένα

1. Εισαγωγή στην Επιδημιολογία και την Δημόσια Υγεία
 - Μέτρα συχνότητας νοσημάτων
 - Επίπτωση
 - Επιπολασμός
 - Είδη επιδημιολογικών Μελετών
 - Συγχρονικές Μελέτες – Μελέτες Επιπολασμού
 - Μελέτες κοορτής – Διαχρονικές Μελέτες
 - Μελέτες Ασθενών - Μαρτύρων
2. Αξιολόγηση διαγνωστικής ικανότητας
 - Αξιοπιστία και Εγκυρότητα
 - Ευαισθησία - Ειδικότητα
 - Καμπύλες ROC
3. Εισαγωγή στις Κλινικές Μελέτες
 - Είδη Κλινικών Μελετών
 - Φάσεις 1 – 4
 - Σχεδιασμός Κλινικών Μελετών
 - Στατιστικές Τεχνικές που χρησιμοποιούνται στις Κλινικές Δοκιμές
4. Δειγματοληψία
5. Ανάλυση Κατηγορικών Δεδομένων
 - Σύγκριση μεταξύ ομάδων
 - Ανάλυση Επαναλαμβανόμενων μετρήσεων
6. Πολυμεταβλητή Ανάλυση
7. Καταληκτικά Σημεία (Endpoints)
8. Αναφορά Αποτελεσμάτων (The CONSORT statement)
9. Πρωτόκολλο
10. Ηθικά Ζητήματα (Good Clinical Practice-GCP)
11. Μοντέλα Ανάλυσης Επιβίωσης
 - Πίνακες επιβίωσης
 - Καμπύλες επιβίωσης Kaplan-Meier και σύγκριση
 - Μοντέλα παλινδρόμησης Cox
12. Συγγραφή και κριτική ενός επιστημονικού άρθρου