



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εδνικόν και Καποδιστριακόν
Πανεπιστήμιον Αδηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

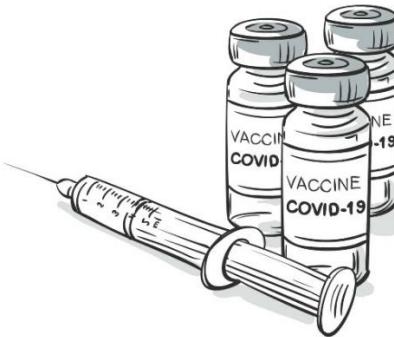
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εδνικό και Καποδιστριακό
Πανεπιστήμιο Αδηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ
ΣΠΟΥΔΩΝ «ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ-
ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΛΟΓΙΑ»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Πρόβλεψη της σοβαρότητας των παρενεργειών των εμβολίων κατά της COVID-19 με χρήση αλγορίθμων μηχανικής μάθησης»



Ελλη Ραμμένου

Πτυχιούχος Τμήματος Βιολογίας, Πανεπιστήμιο Πατρών

ΑΘΗΝΑ 2024



HELLENIC REPUBLIC

National and Kapodistrian
University of Athens

EST. 1837

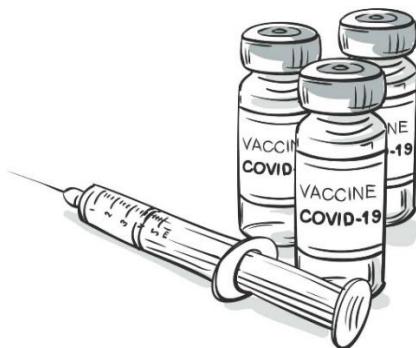
HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

SCHOOL OF SCIENCE
DEPARTMENT OF BIOLOGY

MASTER IN «BIOINFORMATICS-
COMPUTATIONAL BIOLOGY»

Master Diploma Thesis

**«Prediction of severity of COVID-19 vaccines using
machine learning algorithms»**



ELLI RAMMENOU

BSc Biology, University of Patras

ATHENS 2024



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εδνικόν και Καποδιστριακόν
Πανεπιστήμιον Αδηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

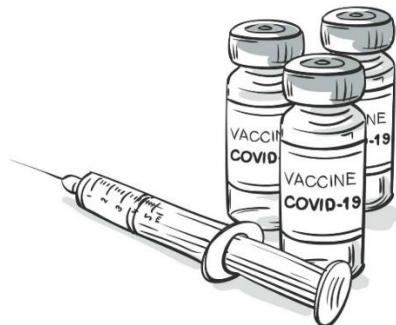
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εδνικό και Καποδιστριακό
Πανεπιστήμιο Αδηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ
ΣΠΟΥΔΩΝ «ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ-
ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΛΟΓΙΑ»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Πρόβλεψη της σοβαρότητας των παρενεργειών των εμβολίων
κατά της COVID-19 με χρήση αλγορίθμων μηχανικής μάθησης»



Τριμελής εξεταστική επιτροπή

Αναπληρώτρια Καδηγήτρια Βασιλική Οικονομίδου (Επιβλέπουσα)

Τομέας Βιολογίας Κυττάρου & Βιοφυσικής

Τμήμα Βιολογίας, ΕΚΠΑ

Καδηγητής Ιωάννης Τρουγκάκος

Τομέας Βιολογίας Κυττάρου & Βιοφυσικής

Τμήμα Βιολογίας, ΕΚΠΑ

Ιωάννης Μιχαλόπουλος, Ειδικός Λειτουργικός Επιστήμονας Β'

Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αδηνών

Ευχαριστίες

Θα ήθελα να απευθύνω ευχαριστίες στους ανθρώπους που με βοήθησαν και με στήριξαν κατά τη διάρκεια της εκπόνησης της συγκεκριμένης διπλωματικής εργασίας.

Αρχικά, θα ήθελα να ευχαριστήσω θερμά τον Ειδικό Λειτουργικό Επιστήμονα Β' του Ιδρύματος Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών, Δρ. Ιωάννη Μιχαλόπουλο, μέλος της εξεταστικής μου επιτροπής που μου προσέφερε μία θέση στο εργαστήριό του και μου παρείχε τα απαραίτητα εφόδια για την πορεία μου ως βιοπληροφορικός, καθώς και χρήσιμες συμβουλές σχετικά με τη γενικότερη επιστημονική προσέγγιση. Θα ήθελα επίσης να απευθύνω τις ευχαριστίες μου στην Αναπληρώτρια Καθηγήτρια Βασιλική Οικονομίδου του Τμήματος Βιολογίας του Καποδιστριακού Πανεπιστημίου Αθηνών, στον Καθηγητή Ιωάννη Τρουγκάκο του Τμήματος Βιολογίας του Καποδιστριακού Πανεπιστημίου Αθηνών, επίσης μέλη της τριμελούς μου εξεταστικής επιτροπής, οι οποίοι υπήρξαν καθηγητές μου κατά τη διάρκεια του Προγράμματος Μεταπτυχιακών Σπουδών μεταφέροντας μου χρήσιμες πληροφορίες και γνώσεις για ποικίλα αντικείμενα του κλάδου της Βιοπληροφορικής.

Θα ήθελα επίσης να ευχαριστήσω το φίλο και συνεργάτη μου Δημήτριο Γεωργίου για την πολύτιμη βιόθειά του στο κομμάτι της μηχανικής μάθησης και της κατασκευής του μοντέλου, τον Βασίλειο Ζωγόπουλο για την πολύτιμη βιόθειά του στο τεχνικό κομμάτι της εργασίας, καθώς και τους υπόλοιπους συμφοιτητές μου στο εργαστήριο του Δρ. Ιωάννη Μιχαλόπουλου για το ευχάριστο κλίμα συνεργασίας.

Είμαι ιδιαίτερα ευγνώμων στους αγαπημένους μου φίλους και τους γονείς μου που ήταν δίπλα μου όλο αυτό το διάστημα και με στήριξαν στην απόφασή μου να ακολουθήσω αυτόν τον δρόμο.

Περιεχόμενα

1	Εισαγωγή.....	3
1.1	Iοί.....	3
1.1.1	Γενικές πληροφορίες.....	3
1.1.2	Ιστορική αναδρομή	3
1.1.3	Ορισμός.....	4
1.1.4	Δομή	4
1.1.5	Ταξινόμηση	4
1.2	Κορωνοϊοί.....	9
1.2.1	Ιστορική αναδρομή	9
1.2.2	Ταξινομική και γενικά χαρακτηριστικά κορονοϊών	10
1.2.3	Αλληλεπίδραση ιού – ξενιστή	11
1.2.4	SARS - CoV	12
1.2.5	SARS – CoV – 2	13
1.3	Ασθένεια CoViD-19.....	17
1.4	Εμβόλια κατά της CoViD-19	18
1.4.1	Εμβόλιο των Pfizer–BioNTech mRNA BNT162b2	18
1.4.2	Εμβόλιο της Moderna (mRNA-1273)	23
1.4.3	Εμβόλιο της AstraZeneca	24
1.4.4	Εμβόλιο της Johnson & Johnson (Janssen)	25
1.4.5	Σύνοψη χαρακτηριστικών εμβολίων	28
1.5	Βάση Δεδομένων EudraVigilance	29
1.5.1	Εισαγωγή	29
1.5.2	Δεδομένα της EudraVigilance	29
1.5.3	Αναζήτηση στη βάση.....	33
1.6	Μηχανική μάθηση.....	38
1.6.1	Ορισμός.....	38
1.6.2	Διαχωρισμός συνόλου δεδομένων	39
1.6.3	Τύποι Μηχανικής Μάθησης.....	40
1.6.4	SHAP (SHapley Additive exPlanations).....	42
2	Σκοπός.....	44
3	Μέθοδοι.....	45
3.1	Σύνολο δεδομένων – Γενικά Χαρακτηριστικά.....	45
3.2	Parsing του συνόλου δεδομένων	46
3.2.1	Γλώσσα PHP	46

3.2.2	Parser	47
3.3	Δημιουργία βάσης δεδομένων	49
3.3.1	Γλώσσα SQL.....	50
3.3.2	MySQL DBMS.....	51
3.3.3	MySQL Workbench.....	53
3.3.4	Διάγραμμα οντοτήτων - συσχετίσεων (ERD)	54
3.4	Μηχανικη μαθηση.....	59
3.4.1	Δημιουργία αρχείου input για την εκπαίδευση του μοντέλου	59
3.4.2	Γλώσσα Python.....	60
3.4.3	Data pre-processing (προ-επεξεργασία δεδομένων)	61
3.4.4	Διαχωρισμός συνόλου δεδομένων	65
3.4.5	Αλγόριθμοι κατηγοριοποίησης	65
3.4.6	SHAP (SHapley Additive exPlanations).....	76
3.5	Μετρικές αξιολόγησης	77
3.5.1	Accuracy	79
3.5.2	Recall	80
3.5.3	Precision	80
3.5.4	F1 score	81
3.6	Γενικές τεχνικές πληροφορίες	82
4	Αποτελέσματα	83
4.1	Περιγραφή συνόλου δεδομένων	83
4.1.1	Συνολικός αριθμός εγγραφών	83
4.1.2	Περιγραφικά στοιχεία	83
4.2	Διάγραμμα οντοτήτων – συσχετίσεων.....	88
4.3	Προ-επεξεργασία δεδομένων	91
4.4	Κλάση του μοντέλου ταξινόμησης	91
4.5	Μετρικές αξιολόγησης αλγορίθμων ταξινόμησης	94
4.5.1	Πείραμα A	95
4.5.2	Πείραμα B	95
4.5.3	Πείραμα Γ	96
4.6	Αποτελέσματα ανάλυσης SHAP	101
5	Συζήτηση	105
5.1	Μορφή του συνόλου δεδομένων.....	105
5.2	Τεχνικές προ-επεξεργασίας δεδομένων	106
5.3	Απόδοση αλγορίθμων ταξινόμησης.....	108
5.3.1	Σύγκριση για διαφορετικά μεγέθη συνόλου δεδομένων	108

5.3.2	Συνολική επίδραση του μεγέθους του δείγματος.....	110
5.4	Ανάλυση SHAP.....	111
5.5	Εναλλακτικές Προσεγγίσεις.....	114
6	Συμπεράσματα.....	115
7	Βιβλιογραφία	116