**Πρόταση πτυχιακής εργασίας για το M.Sc. Data Science   
(Athens Metropolitan College)**

# Βασικές πληροφορίες

**Ημερομηνία**: 17/5/2021

**Ονοματεπώνυμο**: Νίκος Κουγιανός **UEL Number**: 2018506

**Προτεινόμενος τίτλος (αγγλικά)**   
Machine learning popular algorithms: In-depth analysis and hands-on implementation.

**Προτεινόμενος τίτλος (ελληνικά)**Δημοφιλείς αλγόριθμοι μηχανικής εκμάθησης: Εκτεταμένη ανάλυση και πρακτικές υλοποίησης.

# Πρόταση

## Θεωρητικό μέρος

Στα πλαίσια του θεωρητικού μέρους και της βιβλιογραφικής έρευνας της πτυχιακής εργασίας θα γίνει εκτενής αναφορά σε κάποιους από τους δημοφιλέστερους αλγορίθμους μηχανικής εκμάθησης, η οποία θα βασίζεται στους παρακάτω πυλώνες, για τον κάθε αλγόριθμο:

* Ορισμός και βασικές πληροφορίες
* Τρόπος λειτουργίας
* Πλεονεκτήματα και περιορισμοί

Στη συνέχεια, χρησιμοποιώντας ως σημείο αναφοράς ερευνητικές δουλειές και δημοσιεύσεις / επιστημονικά άρθρα καταξιωμένων data scientists (και όχι μόνο), θα γίνει σύγκριση των επιλεχθέντων αλγορίθμων, ως προς την ακρίβεια, την ευαισθησία, τον αριθμό απαιτούμενων attributes κλπ, προκειμένου να εξαχθούν γενικότερα συμπεράσματα για τα δυνατά σημεία και τις αδυναμίες του καθενός.

Οι αλγόριθμοι που έχουν επιλεχθεί μέχρι στιγμής ανήκουν στην οικογένεια της εποπτευόμενης μάθησης (supervised learning) και είναι οι παρακάτω:

* Logistic/Linear Regression
* Naive Bayes
* Support Vector Machine
* Decision Tree

Σαφώς, αναλόγως του διαθέσιμου χρόνου και της έκτασης που θα καταλάβει η παραπάνω έρευνα, και πάντα σε συνεννόηση με τους επιβλέποντες καθηγητές, ενδέχεται να προστεθεί αντίστοιχη ενότητα και για αλγορίθμους μη εποπτευόμενης μάθησης (unsupervised learning):

* K-Means
* KNN

Παρ’ όλα αυτά, αποτελεί ξεκάθαρη επιθυμία μου η αφιέρωση του μεγαλύτερου μέρους του χρόνου, στο πρακτικό κομμάτι της πτυχιακής εργασίας.

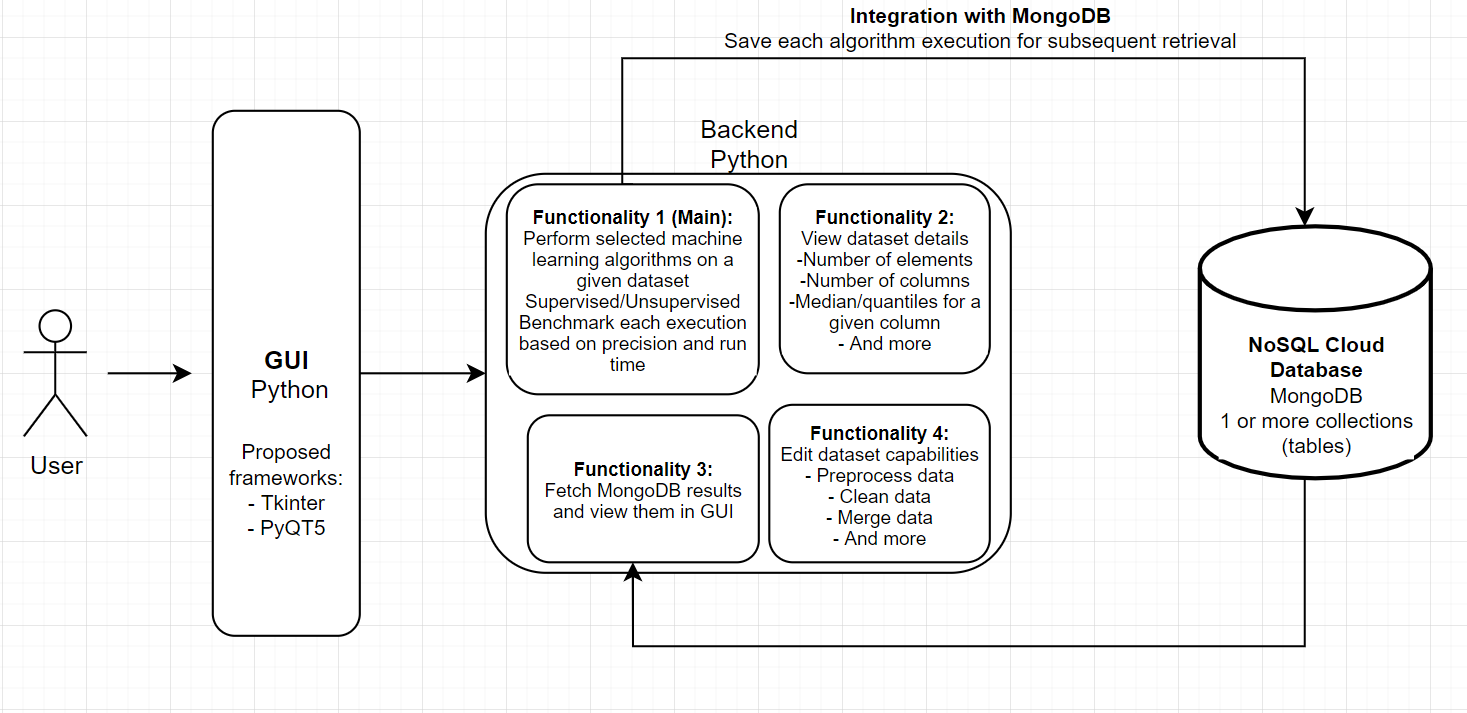
## Πρακτικό μέρος

### Αρχική σκέψη

Στα πλαίσια του πρακτικού μέρους της εργασίας, θα γίνει απόπειρα υλοποίησης ολοκληρωμένης desktop εφαρμογής, η οποία εν συντομία θα προσπαθήσει να δώσει στον χρήστη έναν εύχρηστο τρόπο διαχείρισης ενός συνόλου δεδομένων, θα εφαρμόζει συγκεκριμένους αλγορίθμους μηχανικής εκμάθησης σε κάποιο dataset, και θα προσφέρει και κάποιες επιπλέον δυνατότητες όπως διασύνδεση και αποθήκευση δεδομένων σε κάποια απομακρυσμένη βάση δεδομένων.

### Προτεινόμενη αρχιτεκτονική

Η εφαρμογή θα ακολουθήσει το κλασικό μοτίβο που ακολουθούν όλα τα μοντέρνα web/desktop applications χωρίζοντας το project σε ένα Γραφικό Περιβάλλον Χρήστη (GUI) και σε ένα Backend κομμάτι το οποίο θα αναλαμβάνει την διαχείριση εντολών και του input που θα δίνει ο χρήστης. Μέσω του Backend θα γίνει διασύνδεση και σε μια Cloud NoSQL βάση δεδομένων (MongoDB) η οποία αναλύεται παρακάτω. Το προτεινόμενο αρχιτεκτονικό διάγραμμα είναι το εξής:



### Frontend

Για τεχνολογία frontend, δηλαδή το εργαλείο με το οποίο θα φτιαχτεί όλο το γραφικό περιβάλλον, προτείνεται το Python framework PyQT5, και ως δεύτερη επιλογή το Tkinter. Περαιτέρω πληροφορίες και για τα 2 βρίσκονται στο footnote[[1]](#footnote-1). Εδώ να σημειωθεί ότι δεν υπάρχει σχετική εμπειρία και εξοικείωση με κανένα framework που να φτιάχνει γραφικά desktop περιβάλλοντα, οπότε ό,τι επιλεχθεί θα είναι από το 0. Η επαγγελματική μου εμπειρία και οι εις βάθος γνώσεις μου είναι Java Backend development (Spring Boot), και μισός χρόνος σε FullStack Web development (PHP). Συνεπώς, κάθε πρόταση στο συγκεκριμένο κομμάτι είναι ευπρόσδεκτη, αν και γενικότερα από την έρευνα που έχει γίνει το PyQT έχει αρκετό support και δυνατό developer community και παράγει άρτια αποτελέσματα.

Όσον αφορά τα design mockups, δηλαδή τις οθόνες που θα βλέπει ο end user που θα αλληλεπιδρά με το πρόγραμμα, αυτά θα φτιαχτούν στην φάση της ανάλυσης/design του πρότζεκτ (βλ. πλάνο παρακάτω)

### Backend

### Διασύνδεση με βάση δεδομένων

### Επιπλέον λεπτομέρειες / Μελλοντικές ενέργειες

### Λεπτομέρειες άδειας χρήσης

# Πλάνο

1. Tkinter: <https://wiki.python.org/moin/TkInter>

   PyQT5: <https://en.wikipedia.org/wiki/PyQt>

   Σύγκριση: <https://dev.to/amigosmaker/python-gui-pyqt-vs-tkinter-5hdd> [↑](#footnote-ref-1)