

# Ατομική Έκθεση Εργασίας

## Στοιχεία Φοιτητή

Ονοματεπώνυμο	Κρανίτσα Αντωνία
Αριθμός Μητρώου	std168242
Στοιχεία Επικοινωνίας	Email: <a href="mailto:std168242@ac.eap.gr">std168242@ac.eap.gr</a> Κινητό Τηλέφωνο: 6988591742
Ακαδημαϊκό Έτος	2024-2025
Τμήμα	ΗΛΕ52
Ονοματεπώνυμο Καθηγητή Συμβούλου	Νικόλαος Ευθυμιόπουλος
Project ID	40
Τίτλος Project	Πρόβλεψη Καμπάνιας Marketing με χρήση αλγορίθμων Μηχανικής Μάθησης (Machine Learning)

## Πίνακας περιεχομένων

Σκοπός του Project .....	2
Ατομική Συνεισφορά .....	2
Ώρες Εργασίας .....	3
Βιβλιογραφία .....	3

## Σκοπός του Project

Η παρούσα ατομική εργασία υλοποιήθηκε για την ανάπτυξη ενός λογισμικού Μηχανικής Μάθησης με τη χρήση της γλώσσας προγραμματισμού Python, το οποίο ενσωματώνει γραφικό περιβάλλον χρήστη (GUI). Χρησιμοποιεί αλγόριθμο k-NN (k-Nearest Neighbors) μέσω του οποίου προβλέπει την ανταπόκριση πελατών σε προωθητικές ενέργειες μάρκετινγκ απεικονίζοντας κιάλας κάποια αποτελέσματα σε μορφή διαγραμμάτων για κατανόηση δεδομένων και σύγκριση κατηγοριών. Για να το πετύχει αυτό, δέχεται και επεξεργάζεται δεδομένα προηγούμενων καμπανιών που περιλαμβάνουν λίστες με στοιχεία 1000 πελατών και την ανταπόκρισή τους. Σκοπός αυτού του project είναι η κατανόηση κα πρόβλεψη της συμπεριφοράς των νέων πελατών με βάση τα προηγούμενα δεδομένα και στοιχεία τους.

## Ατομική Συνεισφορά

Αυτή η αναφορά εστιάζει στη δική μου συμμετοχή και συνεισφορά στην υλοποίηση του συγκεκριμένου project. Αρχικά σε συνεργασία με τον Πίτσαρη Κωνσταντίνο δημιουργήσαμε το βασικό παράθυρο της εφαρμογής (το root window) ρυθμίζοντας και άλλες παραμέτρους όπως το θέμα της μέσω της βιβλιοθήκης `sv_ttk`, τον τίτλο και τις διαστάσεις του παραθύρου. Επίσης φτιάξαμε τα μέλη της κλάσης τα οποία θα δεχτούν τα απαραίτητα δεδομένα υλοποίησης του προγράμματος όπως `past_campaign_data` (δεδομένα προηγούμενης καμπάνιας), `new_campaign_data` (δεδομένα νέας καμπάνιας), `knn_model` (το μοντέλο Μηχανικής Μάθησης (k-Nearest Neighbors)) και `predictions_df` (DataFrame με τα αποτελέσματα των προβλέψεων) αλλά και τις σημαίες κατάστασης (flags) που χρησιμοποιούνται για να ελέγχουν την πρόοδο/ροή της εφαρμογής για φόρτωση παλιών και νέων δεδομένων, εκπαίδευση μοντέλου και για τις προβλέψεις. Επίσης δημιούργησα μια χειροκίνητη εκπαίδευση του μοντέλου K-NN με τιμή k που εισάγει ο χρήστης όπου η μέθοδος ζητάει από τον χρήστη έναν ακέραιο k και έτσι δημιουργεί νέο KNN αντικείμενο. Κάνει δηλαδή έναρξη της διαδικασίας εκπαίδευσης του μοντέλου τροφοδοτώντας τα δεδομένα εκπαίδευσης και εκπαιδεύοντας το μοντέλο πρόβλεψης. Ελέγχει αν έχουν φορτωθεί τα δεδομένα αυτά και δημιουργεί το αντικείμενο KNN. Καταγράφει και μεταφέρει τα δεδομένα ενώ αναζητά τον βέλτιστο αριθμό γειτόνων (k)

με cross-validation. Τέλος ενημερώνει τα flags κατάστασης και τα κουμπιά της διεπαφής ανάλογα με την έκβαση. Στο τέλος του project δημιουργεί και εμφανίζει ένα διάγραμμα πίτας που απεικονίζει τις ανταποκρίσεις των αποτελεσμάτων των πελατών χωριστά ανά φύλο ενώ αν δεν υπάρχουν δεδομένα προβλέψεων ('self.predictions\_df' είναι None), εμφανίζει μήνυμα λάθους και καταγράφει σχετικό μήνυμα στο αρχείο καταγραφής. Μετατρέπει τα ονόματα στηλών σε μικρά γράμματα χωρίς κενά και μετονομάζει τις στήλες 'Φύλο' και 'Ανταπόκριση' σε 'gender' και 'response' αντίστοιχα, ώστε να υπολογίσει το ποσοστό των 'yes' απαντήσεων ανά φύλο και σχεδιάζει δύο διαγράμματα πίτας (ένα για κάθε φύλο) με διαφορετικά χρώματα και τίτλους. Τέλος, ασχολήθηκα εξ ολοκλήρου με τη συγγραφή της παρουσίασης της ομαδικής εργασίας σε μορφή διαφανειών.

Σε συνεργασία με τα υπόλοιπα μέλη της ομάδας, συνέβαλα στη δομή του κώδικα, εντοπισμό και διόρθωση σφαλμάτων κώδικα, διαμόρφωση της γενικότερης αισθητικής της εφαρμογής, συγγραφή της ομαδικής εργασίας του project ενώ εκτέλεσα τη συγγραφή της ατομικής έκθεσης εργασιών.

## **Ώρες Εργασίας**

Μία πρόχειρη εκτίμηση (καθώς μία ακριβής ήταν ανέφικτο να γίνει κατά τη διάρκεια) των ωρών που χρειάστηκαν για την εκπόνηση του project είναι:

- περίπου 12 ώρες αναζήτησης βιβλιογραφίας
- περίπου 30 ώρες για ανάπτυξη κώδικα και αναθεωρήσεις
- περίπου 5 ώρες για εντοπισμό και διόρθωση σφαλμάτων κώδικα
- περίπου 4 ώρες για τη συμβολή στη συγγραφή της ομαδικής εργασίας του έργου
- περίπου 5 ώρες για τη συγγραφή ατομικής έκθεσης εργασίας

## **Βιβλιογραφία**

### **Ιστοσελίδες:**

<https://www.ibm.com/think/topics/machine-learning>

<https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>

<https://www.datacamp.com/blog/what-is-machine-learning>

<https://www.elastic.co/what-is/knn>

<https://www.geeksforgeeks.org/k-nearest-neighbours/>

[https://www.w3schools.com/python/python\\_ml\\_knn.asp](https://www.w3schools.com/python/python_ml_knn.asp)

<https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn>

<https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>

<https://docs.python.org/3/library/tkinter.ttk.html>

<https://pythonprogramming.net/styling-gui-bit-using-ttk/>

<https://www.pythontutorial.net/tkinter/ttk-elements/>

<https://www.geeksforgeeks.org/plot-a-pie-chart-in-python-using-matplotlib/>

<https://www.datacamp.com/tutorial/python-pie-chart>

<https://www.analyticsvidhya.com/blog/2024/02/pie-chart-matplotlib/>

### **Videos:**

<https://youtu.be/MPiz50TsyF0?si=2vU0BwihZC3ag9bg>

<https://youtu.be/cJte-7cikzc?si=0JgjlCu6i39tcJFH>

<https://youtu.be/GOuUGWGUT14?si=mVo8NyGY5X7egEBb>

<https://youtu.be/Ecexdzy5h0c?si=Ln4lsTKkF6WjuQg0>

[https://youtu.be/0tM-l\\_ZsxjU?si=jfbkNV6IA\\_CunjDC](https://youtu.be/0tM-l_ZsxjU?si=jfbkNV6IA_CunjDC)

<https://youtu.be/XGWCdEEYp0o?si=x18YrHaVgPy9SPJE>

<https://youtu.be/w6bOBZX-1kY?si=VsMZtiPgOKLekAQm>

[https://youtu.be/gGdlYicGIllo?si=cPj\\_D3WNCw8gmKIY](https://youtu.be/gGdlYicGIllo?si=cPj_D3WNCw8gmKIY)

[https://youtu.be/CQveSaMyEwM?si=\\_op1zCOAPOeWee84](https://youtu.be/CQveSaMyEwM?si=_op1zCOAPOeWee84)

**Σημειώσεις μαθημάτων:**

Διαφάνειες των ΟΣΣ3 και ΟΣΣ4 της ΠΛΗΠΡΟ