



Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών-Μηχανικών  
Ηλεκτρονικών Υπολογιστών  
Διατμηματικό Μεταπτυχιακό Πρόγραμμα  
Μηχανικής Μάθησης και Επιστήμης Δεδομένων

Εξόρυξη Γνώσης από Δεδομένα

Chicago Crime Rate

Μύρων Σαμψάκης-Μπακόπουλος  
Δημήτρης Λαμπράκης  
Σπύρος Πούρος

Αθήνα, Φεβρουάριος 2020

## Περιεχόμενα

1	Εισαγωγή	2
2	Απλά στατιστικά	2
3	Απεικόνιση σε χάρτη	2
4	Δημιουργία νέων χαρακτηριστικών	2
5	Εφαρμογή μεθόδων Μηχανικής Μάθησης	2
6	Απεικόνιση προτύπων	2
7	Σύγκριση συμπερασμάτων με πραγματικά γεγονότα	2
8	Τελικά Συμπεράσματα	2

# 1 Εισαγωγή

Σκοπός της ερασίας αυτής είναι η εφαρμογή τεχνικών εξόρυξης γνώσης από το dataset **Chicago Crime Data** που διατίθεται από το Google Cloud Platform της Google.

Το dataset αυτό αποτελείται από δηλώσεις εγκλημάτων που έγιναν στην περιοχή του Σικάγου κατά περίπου τα τελευταία 20 χρόνια. Πέραν του είδους του εγκλήματος, περιλαμβάνει πληροφορίες όπως η ώρα και η μέρα που έγινε, η περιοχή, εάν υπήρξε σύλληψη ή όχι, ο αριθμός της υπόθεσης, και διάφορα άλλα. Οι συνολικές καταχωρήσεις ανέρχονται περίπου στις 7 εκατομμύρια, καθιστώντας το dataset αρκετά μεγάλο.

Η όλη επεξεργασία θα γίνει μέσω της πλατφόρμας της Google, Google Cloud Platform, με την χρήση queries σε SQL γλώσσα. Θα γίνει αναζήτηση διαφόρων patterns στα δεδομένα (data exploration), θα εξαχθούν μετα-χαρακτηριστικά (χαρακτηριστικά που δεν υπάρχουν στα αρχικά δεδομένα), θα εφαρμοστούν ορισμένες μέθοδοι Machine Learning σε ένα κομμάτι των δεδομένων, με σκοπό το classification, prediction, και clustering τους. Επίσης, θα γίνει ταύτιση των συμπερασμάτων με τα αίτια που πραγματικά τα κίνησαν (νομοθεσίες, οικονομική κρίση, κινήματα ρατσισμού κ.λ.π.). Τέλος, ένα μεγάλο μέρος των δεδομένων θα απεικονιστεί σε διαγράμματα αλλά και σε χάρτες, με σκοπό την διευκόλυνση κατανόησης των συμπερασμάτων.

## 2 Απλά στατιστικά

## 3 Απεικόνιση σε χάρτη

## 4 Δημιουργία νέων χαρακτηριστικών

## 5 Εφαρμογή μεθόδων Μηχανικής Μάθησης

## 6 Απεικόνιση προτύπων

## 7 Σύγκριση συμπερασμάτων με πραγματικά γεγονότα

## 8 Τελικά Συμπεράσματα