

# ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΣΤΗΝ ΕΚΤΙΜΗΣΗ ΤΗΣ ΑΚΙΝΗΤΗΣ ΠΕΡΙΟΥΣΙΑΣ

## Από την προσωπική εκτίμηση στην αυτόματη αξιολόγηση

**Spyridon Savvidis**

Senior Valuation specialist Allianz Global Investors

Η ακίνητη περιουσία είναι η μεγαλύτερη κατηγορία περιουσιακών στοιχείων (asset class) στον κόσμο. Έλαβε την παγκόσμια βιομηχανική ταξινόμηση τον Αύγουστο του 2016 και αντιπροσωπεύει, κατά μέσο όρο, το 5,1% κάθε θεσμικού χαρτοφυλακίου (Andonov, Eichholtz και Kok [2013]). Ο προσδιορισμός της αξίας των ακινήτων παραμένει εξαιρετικά δύσκολο, ακόμα και για ένα εργατικό δυναμικό των 74.000 εκτιμητών στις Ηνωμένες Πολιτείες, οι οποίοι εξακολουθούν να αξιολογούν με αρκετά υποκειμενικό τρόπο την αξία τους, που μερικές φορές αξίζουν δισεκατομμύρια δολάρια.

Οι εκτιμήσεις βασίζονται συνήθως στην κεφαλαιοποίηση του καθαρού εισοδήματος ενός ακινήτου χρησιμοποιώντας την απόδοση (ή ανώτατη τιμή) που έχει συναχθεί από μεταβιβάσεις κοντινών, παρόμοιων ακινήτων ή από συμφητισμούς αυτών. Ωστόσο, το δείγμα αυτών των ακινήτων δεν είναι ποτέ πραγματικά συγκρίσιμο, ούτε με το χρόνο πώλησης, ούτε με τα χαρακτηριστικά του ακινήτου και συνήθως συνδέονται με παλιές αποτιμήσεις ή με την αντικειμενική αξία του. Το αποτέλεσμα είναι ότι οι εκτιμήσεις ακινήτων συνήθως δεν αντικατοπτρίζουν την υπάρχουσα κατάσταση στην αγορά και παρέχουν ένα μέσο όρο των προσεγγίσεων των πραγματικών αγοραίων τιμών, με τιμές τεχνητά χαμηλές στις ανοδικές αγορές και υψηλές στις πτωτικές αγορές.

Αυτά τα αβέβαια μέτρα υπολογισμού αξίας είναι προφανώς προβληματικά στην σημερινή, καθ' άλλως εξαιρετικά αποτελεσματική κεφαλαιαγορά: Οι δείκτες τιμών ακινήτων που βασίζονται στις εκτιμήσεις, οι οποίες εξακολουθούν να χρησιμοποιούνται από πολλούς επενδυτές ως το σημείο αναφοράς επιλογής, είναι πολλές φορές εσφαλμένες. Οι δανειστές και οι επενδυτές δεν έχουν πρόσβαση σε ακριβή στοιχεία τιμολόγησης για την υποθήκευση ακινήτων. Η αξιολόγηση του κινδύνου ακίνητης περιουσίας στους ισολογισμούς των τραπεζών και των επενδυτών είναι ακόμα δύσκολη, κάτι που φάνηκε ιδιαίτερα κατά την τελευταία συγκυρία το 2007-2012.

Η ακρίβεια των εκτιμήσεων υπήρξε θέμα δημοφιλών συζητήσεων και ακαδημαϊκών σπουδών για πολλές δεκαετίες. Πρόσφατα, η Cannon and Cole [2011] ανέλυσε την ακρίβεια των εκτιμήσεων για τον τομέα εμπορικών ακινήτων των ΗΠΑ, χρησιμοποιώντας τα στοιχεία του 1984-2010 από το Εθνικό Συμβούλιο των Καταπιστευματικών Επενδύσεων Ακινήτων (NCREIF). Συγκρίνοντας τις εκτιμήσεις ακινήτων με τις πραγματικές συναλλαγές, οι συντάκτες τεκμηρίωσαν ότι κατά μέσο όρο οι εκτιμήσεις κατά 12% υπερβαίνουν πάνω ή κάτω από την επόμενη τιμή συναλλαγής (διόρθωση για την χρονική καθυστέρηση μεταξύ συναλλαγών ακινήτων και αποτιμήσεων). Τα αποτελέσματα αυτά συμφωνούν με τα ευρήματα των Fisher, Miles και Webb [1999] για την περίοδο 1978-1998. Τόνισαν μια μέση απόλυτη απόκλιση από 9% σε 12,5% μεταξύ των εκτιμήσεων και των τιμών των συναλλαγών.

Εκτός των Ηνωμένων Πολιτειών, η MSCI δημοσίευσε μια ετήσια έκθεση έρευνας που για περισσότερο από μια δεκαετία σύγκρινε τις αποτιμήσεις και τις τιμές πώλησης. Η έκθεση MSCI Ιουνίου 2016 παρουσιάζει σημαντικές διαφορές μεταξύ των χωρών και των ετών, αλλά η απόλυτη διαφορά μεταξύ εκτιμήσεων ακινήτων και πραγματικών συναλλαγών κυμαίνεται μεταξύ 7,7% (Ιταλία) και 13,9% (Ιαπωνία). Παρόλο που η ακρίβεια και η αναποτελεσματικότητα των εκτιμήσεων ακινήτων παραμένουν, ο τομέας των ακινήτων έχει σημειώσει σημαντική αύξηση στη διαθεσιμότητα δεδομένων και την εμφάνιση των τεχνικών μηχανικής μάθησης, οι οποίες χρησιμοποιούνται σήμερα ευρέως στην ιατρική έρευνα και σε εφαρμογές όπως αλγόριθμοι αναζήτησης και μηχανισμών συστάσεων (παραδείγματα εφαρμογών μηχανικής μάθησης ποικίλλουν σε μεγάλο βαθμό, από την ανίχνευση απάτης από την PayPal έως την εξατομικευμένη διαφήμιση στο διαδίκτυο που έχουμε συνηθίσει). Στην αγορά κατοικίας, μερικά νεοσυσταθέντα προϊόντα χρησιμοποιούν το συνδυασμό αφθονίας δεδομένων και μοντέλων μηχανικής μάθησης, συμπεριλαμβανομένου, παραδείγματος χάριν, του

αυτοματοποιημένου μοντέλου αποτίμησης του House Canary και του πολυσυζητημένου "Zestimate" της Zillow.

Σε αυτή τη μελέτη, παρέχουμε μια πρακτική εφαρμογή μεγάλων δεδομένων σε συνδυασμό με εξελιγμένες τεχνικές μοντελοποίησης για την ανάπτυξη ενός αυτοματοποιημένου, μηχανοκίνητου μοντέλου αποτίμησης για τον τομέα των εμπορικών ακινήτων. Επικεντρωνόμαστε στον τομέα των κατοικιών, ο οποίος επιτρέπει την πρόσβαση σε ένα σύνολο δεδομένων περίπου 300.000 ακινήτων στην Ελλάδα. Αντί τα παραδοσιακά hedonic μοντέλα, τα οποία περιορίζονται τόσο στατιστικά όσο και από την προδιάθεση του ερευνητή προς τις "συνήθειες" επεξηγηματικές μεταβλητές, εφαρμόζουμε στη συνέχεια μοντέλα υποβοηθούμενης μηχανικής μάθησης που βασίζονται σε (στοχαστικά) δέντρα απόφασης. Αυτά τα μοντέλα μπορούν να "τρέξουν" εκατομμύρια συνδυασμούς χιλιάδων μεταβλητών, να εκπαιδεύσουν και να δοκιμάσουν το μοντέλο σε τυχαία επιλεγμένα μέρη των συνόλων δεδομένων, οδηγώντας σε ακριβείς δοκιμές τυχαίων προβλέψεων. Βρήκαμε ισχυρά στοιχεία για την υπεροχή των αυτοματοποιημένων μοντέλων αποτίμησης (AVMs - Automated valuation models) σε σχέση με τις παραδοσιακές εκτιμήσεις - το μέσο απόλυτο σφάλμα του αυτοματοποιημένου μοντέλου που αναπτύξαμε είναι 15%, το οποίο συγκρίνεται ευνοϊκά με την ακρίβεια των παραδοσιακών εκτιμήσεων και το μοντέλο μπορεί να παράγει μια άμεση τιμή σε κάθε στιγμή με πολύ χαμηλό κόστος. Παρέχουμε επίσης στοιχεία σχετικά με τη σημασία της χρήσης τοπικών πληροφοριών σχετικά με τη θέση ενός ακινήτου. Αν και η χρήση οικονομικών και δημογραφικών δεδομένων από τις απογραφές και ανάλυση στο επίπεδο πόλης είναι συνήθειες πρακτικές στον τομέα της μοντελοποίησης, νέες ομάδες στοιχείων συλλέγονται από μια μεγάλη ποικιλία πηγών, συμπεριλαμβανομένων των δεδομένων από μικρές αγγελίες, των δεδομένων που δίνονται από το κράτος και των ανέσεων που σχετίζονται με την οικονομική δυναμική της περιοχής, προσθέτουν σημαντική αξία στα μοντέλα εκτίμησης.

Η μελέτη αυτή είναι σημαντική για πολλούς λόγους. Πρώτον, ακριβείς και έγκαιρες εκτιμήσεις των αξιών των ακινήτων είναι κρίσιμες για τους επενδυτές ακινήτων και τους δανειστές να λαμβάνουν ενημερωμένες αποφάσεις αναδοχής. Δεύτερον, όπως επεσήμαναν οι Cannon και Cole [2011], «οι επενδυτές, οι ρυθμιστικές αρχές και άλλοι βασίζονται σε εκτιμημένες αξίες για την αξιολόγηση των αποδόσεων των 11 τρισεκατομμυρίων δολαρίων ΗΠΑ»

Στην αγορά ακινήτων, οι αυτοματοποιημένες αποτιμήσεις ακινήτων μπορούν να παράσχουν μια ακριβέστερη εικόνα τόσο του αποθέματος σε ακίνητη περιουσία (δηλαδή της αξίας των ακινήτων στον ισολογισμό) όσο και των ροών (δηλ. των αποδόσεων των ακινήτων από τις μεταβολές των αξιών κεφαλαίου). Τρίτον, τα συστήματα AVM μπορούν να χρησιμοποιηθούν για προσομοίωση ακραίων καταστάσεων υπό δυσμενείς οικονομικές συνθήκες, οι οποίες παραμένουν ένα πολύ αναγκαίο εργαλείο για τις

ρυθμιστικές αρχές, τις τράπεζες, τους οργανισμούς αξιολόγησης και τους επενδυτές. Τέταρτον, η διαθεσιμότητα μιας άμεσης και ακριβούς αξίας ιδιοκτησίας μπορεί να προωθήσει την καινοτομία στον τομέα των ακινήτων, όπως η αυτοματοποιημένη δημιουργία δανείων από τις τράπεζες, τα προϊόντα καθορισμένων εισφορών για ιδιωτικές επενδύσεις σε ακίνητα και τα προϊόντα αρμπιτράζ για εμπορικά ακίνητα (συγκρίσιμα με τα αναδυόμενα προϊόντα για τον τομέα της οικιστικών ακινήτων, όπως το Opendoor.com).

## ΜΕΘΟΔΟΛΟΓΙΑ

Τα AVM είναι μια μεθοδολογία που χρησιμοποιεί μαθηματικά μοντέλα για να παρέχει την εκτιμώμενη αξία μιας ιδιοκτησίας σε ένα συγκεκριμένο χρονικό σημείο. Τα εργαλεία AVM μπορούν να υιοθετήσουν διαφορετικές προσεγγίσεις για τον υπολογισμό της αξίας μιας ιδιοκτησίας. Είναι σημαντικό ότι το μοντέλο δεν εξαρτάται από τη χρήση ενός ποσοστού απόδοσης/κεφαλαιοποίησης (cap rate), το οποίο είναι κρίσιμο στις παραδοσιακές τεχνικές αποτίμησης ακινήτων. Αυτή η μεταβλητή προκύπτει απλώς από το καθαρό λειτουργικό εισόδημα (Net Operating Income – NOI) ενός περιουσιακού ακινήτου διαιρούμενο με την τιμή συναλλαγής του. Για λόγους αποτίμησης, συνήθως χρησιμοποιούνται οι ανώτατες τιμές των τριών έως πέντε πρόσφατα συναλλασσόμενων κοντινών/σχετικών ακινήτων, προσαρμόζοντας τις διαφορές μεταξύ της εκτιμώμενης αξίας και χαρακτηριστικών του εν λογού ακινήτου και των χαρακτηριστικών των ακινήτων που έχουν μεταβιβαστεί. Αντίθετα, ένα AVM ενσωματώνει όλες τις συναλλαγές σε μια δεδομένη αγορά, υποθέτοντας μια σχέση μεταξύ αξίας και ανεξάρτητων μεταβλητών που είναι παρόμοια μεταξύ των διαφορετικών τοποθεσιών. Τα περισσότερα συστήματα AVM χρησιμοποιούν σήμερα μοντέλα παλινδρόμησης (hedonic) (π.χ. O'Neill [2004], Schulz, Wersing και Werwatz [2014]). Ωστόσο, μερικοί πιο σύγχρονοι αλγόριθμοι χρησιμοποιούν μοντέλα προσαρμοστικής εκτίμησης ή μοντέλα νευρικών δικτύων αυτοδιδασκαλίας (Crosby et al., [2016]).

## HEDONIC ΠΡΟΣΕΓΓΙΣΗ

Hedonic AVM (π.χ. regression models) χρησιμοποιούνται στον τομέα των ακινήτων. Μετά το Rosen [1974], τα hedonic μοντέλα εκτίμησης απαιτούν πληροφορίες για τα βασικά χαρακτηριστικά του ακινήτου, όπως το έτος κατασκευής, το μέγεθος, αριθμός χώρων, καθώς και τα γεωγραφικά χαρακτηριστικά. Αυτά τα μοντέλα συνήθως περιλαμβάνουν μια μηχανή αναζήτησης που συγκρίνει τις ιδιότητες της ιδιότητας του αντικειμένου με συγκρίσιμες ιδιότητες χρησιμοποιώντας ένα πρότυπο αναζήτησης εύρους ή με άλλες λογικές παράμετροι αναζήτησης για μια προκαθορισμένη περίοδο (βλ. Eichholtz, Kok, και Quigley [2010] για μια αίτηση). Τα Hedonic AVMs είναι κοινά επειδή βασίζονται σε απλά μοντέλα παλινδρόμησης που είναι

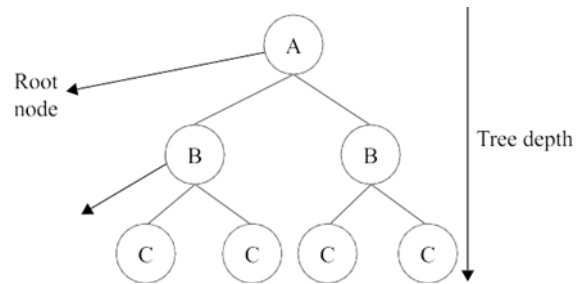
εύκολο να υλοποιηθούν και να κατανοηθούν. Το κύριο μειονέκτημα είναι ότι τα μοντέλα παλινδρόμησης είναι “γενικά” μοντέλα, πράγμα που σημαίνει ότι παράγουν μια ενιαία μέθοδο πρόγνωσης που είναι σταθερή σε όλο το εύρος μεταβλητών (π.χ. ιδιότητες, περιοχή). Ωστόσο, πολλές μεταβλητές έχουν μη γραμμική σχέση σε σχέση με την προβλεπόμενη τιμή. Συνεπώς, ένα μόνο μοντέλο μπορεί να μην είναι η καλύτερη λύση στην πρόβλεψη της αξίας του ακινήτου με τον πιο ακριβή τρόπο.

## Η ΠΡΟΣΕΓΓΙΣΗ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Αντί των παραδοσιακών hedonic μοντέλων, μπορούμε επίσης να εφαρμόσουμε μοντέλα δέντρων αποφάσεων. Τέτοια μοντέλα διαχωρίζουν διαδοχικά το σύνολο δεδομένων σε υποσύνολα για να εφαρμόσουν ένα μοντέλο παλινδρόμησης σε κάθε υποσύνολο. Υπάρχουν δύο τύποι δέντρων αποφάσεων, ανάλογα με τον τύπο της μεταβλητής στόχου: τα ταξινομικά δένδρα (Classification trees), τα οποία αποσκοπούν στην πρόβλεψη κατηγορικών μεταβλητών και τα δέντρα παλινδρόμησης (Regression trees), τα οποία προβλέπουν συνεχείς μεταβλητές (Breiman et al., 1984). Επειδή η αξία ενός ακινήτου είναι συνεχής, χρησιμοποιούμε δέντρα παλινδρόμησης για τα AVM που αναπτύσσονται σε αυτό το άρθρο. Ένας αλγόριθμος δένδρου παλινδρόμησης βρίσκει τους καλύτερους προγνωστικούς δείκτες από το σύνολο ανεξάρτητων μεταβλητών, αρχικά ελαχιστοποιώντας τη διακύμανση της παλινδρόμησης μεταξύ κάθε συνδυασμού εξαρτημένης και ανεξάρτητης μεταβλητής. Αυτό αποδίδει τη σειρά σπουδαιότητας των μεταβλητών. Κατόπιν, κάθε μία από τις επεξηγηματικές μεταβλητές αντιπροσωπεύει έναν κόμβο στο δέντρο αποφάσεων. Ο πρώτος κόμβος, που ονομάζεται επίσης κόμβος ρίζας, περιέχει ολόκληρο το σύνολο δεδομένων (ή τουλάχιστον το τμήμα του συνόλου δεδομένων που χρησιμοποιείται για την εκπαίδευση του μοντέλου). Το σημείο διαίρεσης μέσα στον ριζικό κόμβο βασίζεται στον καλύτερο προγνωστικό παράγοντα και προσδιορίζεται εφαρμόζοντας μια συνάρτηση κόστους που ελαχιστοποιεί τη διακύμανση σε μια γραμμή παλινδρόμησης και για τα δύο υποδείγματα. Η σειρά σπουδαιότητας υπολογίζεται και πάλι για το σύνολο ανεξάρτητων μεταβλητών, με βάση το δείγμα στον κόμβο των φύλλων (το δεύτερο στρώμα κόμβων στο δέντρο). Επιλέγεται η ανεξάρτητη μεταβλητή με τις καλύτερες επιδόσεις και η διάσπαση επαναλαμβάνεται. Αυτή η διαδικασία είναι αναδρομική σε όλους τους κλάδους ενός δέντρου και επαναλαμβάνεται μέχρι να επιτευχθεί ένα όριο (δηλ. Το επιθυμητό βάθος του δέντρου). Η παράσταση 1 δείχνει ένα τυποποιημένο παράδειγμα της δομής ενός δέντρου αποφάσεων. Γενικότερα, για κάθε επεξηγηματική μεταβλητή (ή πρόβλεψη), τα δεδομένα χωρίζονται σε έναν από τους κόμβους ενός δέντρου αποφάσεων, όπου το επίπεδο του κόμβου στο δέντρο βασίζεται στη δύναμη της σχέσης του με την εξαρτημένη μεταβλητή. Για ένα σημείο διαίρεσης  $m$ , που αντιπροσωπεύει ένα υποδειγματότυπο  $R_m$  με παρατηρήσεις  $N_m$  (τιμές της μεταβλητής στόχου), ο κόμβος επιλέγεται κάτω από τις ακόλουθες συνθήκες:

## ΕΚΘΕΣΗ 1

### Μοντέλα μηχανικής μάθησης - Παράδειγμα δέντρου απόφασης με τέσσερις κόμβους



### ΤΥΠΟΙ ΜΟΝΤΕΛΩΝ

Τα δέντρα αποφάσεων (decision trees) έχουν αρκετά πλεονεκτήματα έναντι των μοντέλων hedotic παλινδρόμησης και άλλων τεχνικών μάθησης μηχανών. Για παράδειγμα, τα δέντρα αποφάσεων είναι απλά κατανοητά και ερμηνευτικά και η στατιστική σημασία μπορεί εύκολα να υπολογιστεί. Τα δέντρα απόφασης είναι επίσης σε θέση να χειριστούν μεταβλητές που δείχνουν κατηγορίες (categorical values π.χ. τύπος ακινήτου), οπότε δεν υπάρχει λόγος να δημιουργηθούν από τον χρήστη αριθμητικές μεταβλητές. Τα δέντρα αποφάσεων μπορούν να κάνουν προβλέψεις σε πολύ σύντομο υπολογιστικό χρόνο, ακόμη και με μεγάλες ποσότητες δεδομένων.

Παρά τα πλεονεκτήματα αυτά, τα δέντρα αποφάσεων έχουν ορισμένους περιορισμούς. Πρώτον, μπορούν να έχουν απεριόριστη κατακόρυφη ανάπτυξη, μέχρις ότου οι κόμβοι να έχουν δείγμα μόνο μιας παρατήρησης. Αυτά τα βαθιά δέντρα παράγουν πολύ περίπλοκα μοντέλα που λαμβάνουν όλο το θόρυβο από τα δεδομένα εκπαίδευσης. Αυτό το φαινόμενο, το οποίο ονομάζεται *overfitting*, δημιουργεί μοντέλα που αποφέρουν “κακές” προβλέψεις για δεδομένα που δεν βλέπουν. Δεύτερον, τα βαθιά δέντρα αποφάσεων δεν είναι ευμετάβλητα υπό την έννοια ότι μια μικρή αλλαγή στα δεδομένα εκπαίδευσης μπορεί να οδηγήσει σε μια μεγάλη αλλαγή στο δέντρο και έτσι μια μεγάλη αλλαγή στην τελική πρόβλεψη. Εάν τα δέντρα απόφασης δεν έχουν μεγάλο δείγμα (δηλαδή είναι πάρα πολύ ρηχά), από την άλλη πλευρά, δεν παράγουν ούτε καλά μοντέλα ούτε καλές προβλέψεις. Αυτό το φαινόμενο ονομάζεται *underfitting*.

Για να ξεπεραστούν οι περιορισμοί των μοντέλων δέντρων αποφάσεων, μπορεί κανείς να δημιουργήσει πολλαπλά δέντρα απόφασης και στη συνέχεια να αποκτήσει τη μέση πρόβλεψη. Αυτό ονομάζεται μέθοδος με σύμπλεγμα δέντρων και είναι μια ισχυρή μέθοδος για την επίτευξη καλύτερης προγνωστικής απόδοσης σε σύγκριση με αυτό που θα προέκυπτε μόνο από τα δέντρα αποφάσεων με τις

καλύτερες προβλέψεις. Ένα από τα πιο δημοφιλή μέθοδος με σύμπλεγμα δένδρων είναι το random forest μοντέλο, το οποίο αποτελείται από δείγματα από το αρχικό σύνολο δεδομένων κατάρτισης. Χρησιμοποιώντας το δείγμα εκκίνησης, ένα δέντρο απόφασης είναι πλήρως ανεπτυγμένο έτσι ώστε κάθε κόμβος του να υπολογίζεται επιλέγοντας τυχαίες επεξηγηματικές μεταβλητές. Η όλη διαδικασία, ξεκινώντας από ένα καινούργιο δείγμα, επαναλαμβάνεται πολλές φορές. Αυτή η διαδικασία εξασφαλίζει ότι κάθε δέντρο αποφάσεων που παράγεται θα είναι διαφορετικό και αδιόρατο με τα άλλα. Η τελική πρόβλεψη της μεταβλητής στόχου είναι η σταθμισμένη ψήφος ή ο μέσος όρος των προβλέψεων όλων των δένδρων στη συλλογή. Τα random forests είναι αρκετά καλά στη μείωση της διακύμανσης. Ως εκ τούτου, είναι κατάλληλα να μειώσουν την overfitting που παράγεται από ενιαία, βαθιά δέντρα απόφασης.

Μια άλλη γνωστή μέθοδος του συνόλου είναι η στοχαστική ώθηση (stochastic boosting). Στα μοντέλα gradient boosting, πολλά μικρά δέντρα αποφάσεων χτίζονται διαδοχικά από κομμάτια που απέμειναν από το προηγούμενο δέντρο. Σε κάθε επανάληψη, ένα δέντρο κατασκευάζεται από ένα τυχαίο υποσύνολο του συνόλου δεδομένων (που επιλέγεται χωρίς αντικατάσταση), παράγοντας μια βαθμιαία βελτίωση του μοντέλου. Τελικά, όλα τα μικρά δέντρα στοιβάζονται μαζί ως ένα σταθμισμένο άθροισμα των όρων. Η συνολική ακρίβεια μοντέλου βελτιώνεται προοδευτικά με κάθε πρόσθετο όρο. Λόγω της δομής τους, τα μοντέλα ενίσχυσης αποσκοπούν στη μείωση της μη τυχαίας επιλογής (selection bias) στα δεδομένα. Για τη μοντελοποίηση του AVM σε αυτό το άρθρο, χρησιμοποιούμε τόσο το μοντέλο random forest όσο και δύο μεθόδους ενίσχυσης κλίσης: gradient boosted trees και XGBoost (eXtreme Gradient Boosting είναι μια προηγμένη εφαρμογή του gradient boosting αλγόριθμου.

#### ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΕΠΙΔΟΣΕΩΝ

Η επιστήμη των δεδομένων (data science) προσφέρει διάφορους βασικούς δείκτες απόδοσης (key performance indicators ή KPIs) που επιτρέπουν την ακριβή μέτρηση της απόδοσης ενός μοντέλου. Σε σύγκριση με τη λήψη οικονομικών αποφάσεων, ένα KPI δεν λέει απαραίτητα την αλήθεια και, κατά συνέπεια, είναι επιθυμητό να χρησιμοποιούμε πολλούς δείκτες. Για τα μοντέλα που αναπτύσσονται σε αυτό το άρθρο, αναπτύσσουμε δύο μετρήσεις: επεξηγηματική δύναμη, μετρημένη με την

διακύμανση ή R2, την συνολική ακρίβεια, μετρούμενη με το μέσο τετραγωνικό σφάλμα (mean squared error ή MSE) ή με το μέσο ή μέσο διάγραμμα απόλυτου ποσοστού σφάλματος (mean or median absolute percentage error) (MAPE και MdAPE, αντίστοιχα). Το βέλτιστο μοντέλο μεγιστοποιεί εξηγηθείσα διακύμανση, ενώ βελτιστοποιεί την ακρίβεια. Προφανώς, υπάρχει ένας συνδυασμός μεταξύ αυτών των δύο KPIs δηλαδή ένα ισχυρό μοντέλο δεν είναι απαραίτητως ακριβές και αντίστροφα.

#### REFERENCES

Αυτο το paper είναι μια προσαρμογή του paper Big Data in Real Estate του Nils Kok, Eija-Leena Koronen, Carmen Adriana Martínez-Barbosa, στην Ελληνική πραγματικότητα χρησιμοποιώντας στοιχεία που συλλέγουμε και αναλύουμε στο Roboequity

- [1] Andonov, A., P. Eichholtz, and N. Kok. "A Global Perspective on Pension Fund Investments in Real Estate." *The Journal of Portfolio Management*, Vol. 39, No. 5 (2013), pp. 32-42.
- [2] Breiman, L., J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.
- [3] Cannon, S., and R. Cole. "How Accurate Are Commercial Real Estate Appraisals? Evidence from 25 Years of NCREIF Sales Data." *The Journal of Portfolio Management*, Vol. 35, No. 5 (2011), pp. 68-88.
- [4] Crosby, H., P. Davis, T. Damoulas, and S.A. Jarvis. "A Spatio-Temporal, Gaussian Process Regression, Real-State Price Predictor." In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York: ACM, 2016.
- [5] Eichholtz, P., N. Kok, and J. Quigley. "Doing Well by Doing Good: Green Office Buildings." *The American Economic Review*, Vol. 100, No. 5 (2010), pp. 2492-2509.
- [6] Fisher, J., M. Miles, and B. Webb. "How Reliable Are Commercial Appraisals? Another Look." *Real Estate Finance*, Vol. 16, No. 3 (1999), pp. 9-15.
- [7] *Administration Quarterly*, Vol. 45, No. 3 (2004), pp. 260-268.
- [8] Rosen, S. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, Vol. 82, No. 1 (1974), pp.
- [9] Marcato, G., and T. Key. "Smoothing and Implications for Asset Allocation Choices." *The Journal of Portfolio Management*, Vol. 33, No. 5 (2007), pp. 85-98.
- [10] MSCI. "Private Real Estate: Valuation and Sale Price Comparison." *Research Report*, June 2016.
- [11] O'Neill, J.W. "An Automated Valuation Model for Hotels." *The Cornell Hotel and Restaurant* 34-55.
- [12] Schulz, R., W. Wersing, and A. Werwatz. "Automated Valuation Modelling: A Specification Exercise." *Journal of Property Research*, Vol. 31, No. 2 (2014), pp. 131-153