



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΗΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Εξόρυξη γνώσης από κείμενα τουριστικών εφαρμογών στον κοινωνικό
ιστό με τεχνικές μηχανικής μάθησης**

Ονοματεπώνυμο : Χατζής Βασίλειος

A. M. : Π2016144

Τριμελής εξεταστική επιτροπή

Κερμανίδου Κάτια (επιβλέπουσα)
Έξαρχος Θεμιστοκλής
Κουρουθανάσης Παναγιώτης

Κέρκυρα
2020 – 2021

ΠΕΡΙΛΗΨΗ

Στόχος της παρούσας πτυχιακής εργασίας, είναι η εξόρυξη γνώσης από κείμενα τουριστικών εφαρμογών στον κοινωνικό ιστό με τεχνικές μηχανικής μάθησης. Ο κύριος σκοπός είναι να βρεθεί η χρησιμότητα διαφόρων τεχνικών μηχανικής μάθησης στον τομέα του τουρισμού, και πιο συγκεκριμένα στην πλατφόρμα της Airbnb.

Το φαινόμενο της κοινής οικονομίας καθοδηγείται από την επιθυμία των ανθρώπων για βιωσιμότητα, απόλαυση της δραστηριότητας και οικονομικά οφέλη, το οποίο προκάλεσε αυξανόμενο ενδιαφέρον από ερευνητές και επιχειρήσεις. Η οικονομία κοινής χρήσης επέτρεψε στους χρήστες να ασχολούνται με την πώληση υπηρεσιών μέσω αξιόπιστων διαδικτυακών πλατφορμών όπως η Uber ή η Airbnb.

Το επίκεντρο της συγκεκριμένης πτυχιακής εργασίας είναι η Airbnb, ένας πάροχος διαδικτυακών πλατφορμών peer-to-peer που έχει γίνει ένα από τα πιο επιτυχημένα μοντέλα στην οικονομία κοινής χρήσης. Πρόκειται για μία ηλεκτρονική πλατφόρμα, που φέρνει σε επικοινωνία τον οικοδεσπότη με τον πελάτη, ενώ ταυτοχρόνως είναι υπεύθυνη για τη διεκπεραίωση της διαδικασίας της κράτησης. Μέσω της συγκεκριμένης πλατφόρμας γίνονται οι συναλλαγές, κατά τις οποίες ο επισκέπτης πληρώνει το εκάστοτε ποσό μόλις 24 ώρες μετά την άφιξη του φιλοξενούμενου του, έτσι ώστε να είναι σίγουρο ότι ο τελευταίος παρέλαβε το κατάλυμα, όπως συμφωνήθηκε εξ αρχής. Τα έσοδα της εταιρίας προέρχονται από το ποσοστό που κρατάει εκείνη από κάθε κράτηση που γίνεται.

Στη συγκεκριμένη μελέτη γίνεται μία βιβλιογραφική ανασκόπηση των τεχνικών και πρακτικών που χρησιμοποιούνται για τη βελτιστοποίηση της λειτουργίας της πλατφόρμας της Airbnb μέσω ενός υποτομέα της τεχνητής νοημοσύνης, τη μηχανική μάθηση.

Λέξεις Κλειδιά: Airbnb, Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Εξόρυξη Γνώσης, Τουριστικές Εφαρμογές, Αλγόριθμοι Μηχανικής Μάθησης

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Περίληψη.....	3
Πίνακας Περιεχομένων.....	4
Λίστα Σχημάτων.....	8
Λίστα Πινάκων.....	9
1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΠΛΑΤΦΟΡΜΑ ΤΗΣ AIRBNB.....	10
1.1. Στόχος και περιγραφή της Airbnb.....	10
1.2. Στατιστικά στοιχεία της Airbnb.....	10
1.3. Ιστορικά στοιχεία της Airbnb.....	11
1.4. Λειτουργία της πλατφόρμας.....	12
1.5. Η διαφορά της μίσθωσης σε σχέση με τα ξενοδοχεία και τις επιχειρήσεις καταλυμάτων.....	13
2. ΕΙΣΑΓΩΓΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ.....	15
2.1. Ορισμός της Τεχνητής Νοημοσύνης.....	15
2.2. Ορισμός της Μηχανικής Μάθησης.....	16
2.3. Υπολογιστικές μηχανές και νοημοσύνη.....	17
2.4. Τύποι προβλημάτων και εργασιών.....	17
2.5. Προσεγγίσεις.....	19
2.5.1. Εκμάθηση με δέντρο απόφασης.....	19
2.5.2. Εκμάθηση με κανόνες συσχέτισης.....	19
2.5.3. Τεχνητά νευρωνικά δίκτυα.....	19
2.5.4. Βαθιά μάθηση / Deep learning.....	19
2.5.5. Επαγωγικός λογικός προγραμματισμός.....	20
2.5.6. Μηχανές διανυσμάτων υποστήριξης.....	20
2.5.7. Ομαδοποίηση.....	20
2.5.8. Δίκτυα Bayes.....	21
2.5.9. Ενισχυτική μάθηση.....	21
2.5.10. Εκμάθηση με δέντρο ομοιότητας.....	21
2.5.11. Γενετικοί αλγόριθμοι.....	21
3. ΠΡΟΒΛΕΨΗ ΤΙΜΩΝ.....	23

3.1. Πρόβλεψη τιμών στην Airbnb με χρήση μηχανικής μάθησης και ανάλυσης συναισθημάτων.....	23
3.1.1. Εισαγωγή.....	23
3.1.2. Χαρακτηριστικά και Μέθοδοι.....	23
3.1.3. Αποτελέσματα.....	25
3.1.4. Συμπεράσματα.....	25
3.2. Ανάλυση των τιμών στην Airbnb χρησιμοποιώντας τεχνικές μηχανικής μάθησης.....	26
3.2.1. Εισαγωγή.....	26
3.2.2. Χαρακτηριστικά και Μέθοδοι.....	27
3.2.3. Αποτελέσματα.....	28
3.2.4. Συμπεράσματα.....	30
3.3. Προβλέψεις γειτονιάς και τιμών για τις καταχωρίσεις του Σαν Φρανσίσκο στην Airbnb.....	31
3.3.1. Εισαγωγή.....	31
3.3.2. Χαρακτηριστικά και Μέθοδοι.....	31
3.3.3. Αποτελέσματα.....	33
3.3.4. Συμπεράσματα.....	33
3.4. Πρόβλεψη των τιμών των καταχωρίσεων της Airbnb σε διαφορετικές πόλεις.....	35
3.4.1. Εισαγωγή.....	35
3.4.2. Χαρακτηριστικά και Μέθοδοι.....	35
3.4.3. Αποτελέσματα.....	36
3.4.4. Συμπεράσματα.....	37
3.5. Learning-based μοντέλο πρόβλεψης τιμών για την Airbnb.....	38
3.5.1. Εισαγωγή.....	38
3.5.2. Χαρακτηριστικά και Μέθοδοι.....	39
3.5.3. Αποτελέσματα.....	39
3.5.4. Συμπεράσματα.....	41
3.6. Γενικά συμπεράσματα και μελλοντικές κατευθύνσεις.....	41
4. ΑΝΑΖΗΤΗΣΗ.....	43
4.1. Εφαρμόζοντας βαθιά μάθηση στην αναζήτηση της Airbnb.....	43
4.1.1. Εισαγωγή.....	43

4.1.2. Χαρακτηριστικά και Μέθοδοι.....	44
4.1.3. Αποτελέσματα.....	44
4.1.4. Συμπεράσματα.....	46
4.2. Κατάταξη αναζήτησης και εξατομίκευση στην Airbnb.....	47
4.2.1. Εισαγωγή.....	47
4.2.2. Χαρακτηριστικά και Μέθοδοι.....	47
4.2.3. Αποτελέσματα.....	48
4.2.4. Συμπεράσματα.....	48
4.3. Εκμάθηση μετρήσεων πρόθεσης κράτησης για την αναζήτηση της Airbnb.....	50
4.3.1. Εισαγωγή.....	50
4.3.2. Χαρακτηριστικά και Μέθοδοι.....	50
4.3.3. Αποτελέσματα.....	51
4.3.4. Συμπεράσματα.....	52
4.4. Εξατομίκευση σε πραγματικό χρόνο με τη χρήση ενσωματώσεων για την κατάταξη αναζήτησης στην Airbnb.....	53
4.4.1. Εισαγωγή.....	53
4.4.2. Χαρακτηριστικά και Μέθοδοι.....	53
4.4.3. Αποτελέσματα.....	54
4.4.4. Συμπεράσματα.....	55
4.5. Γενικά συμπεράσματα και μελλοντικές κατευθύνσεις.....	55
5. ΔΥΝΑΜΙΚΗ ΤΙΜΟΛΟΓΗΣΗ.....	57
5.1. Προσαρμοσμένο μοντέλο παλινδρόμησης για δυναμική τιμολόγηση στην Airbnb.....	57
5.1.1. Εισαγωγή.....	57
5.1.2. Χαρακτηριστικά και Μέθοδοι.....	57
5.1.3. Αποτελέσματα.....	58
5.1.4. Συμπεράσματα.....	60
5.2. Διερεύνηση της χρήσης τεχνητής νοημοσύνης στη μεγιστοποίηση των τιμών στον τομέα του τουρισμού: η εφαρμογή της στην περίπτωση της Airbnb στην Κοινότητα της Βαλένθια.....	61
5.2.1. Εισαγωγή.....	61
5.2.2. Χαρακτηριστικά και Μέθοδοι.....	61

5.2.3. Αποτελέσματα.....	63
5.2.4. Συμπεράσματα.....	65
5.3. Γενικά συμπεράσματα και μελλοντικές κατευθύνσεις.....	66
6. ΙΚΑΝΟΠΟΙΗΣΗ ΤΩΝ ΕΠΙΣΚΕΠΤΩΝ.....	67
6.1. Ένα πελατοκεντρικό σύστημα αξιολόγησης για την αξιολόγηση της ικανοποίησης των επισκεπτών της Airbnb με χρήση μηχανικής μάθησης και NLP.....	67
6.1.1. Εισαγωγή.....	67
6.1.2. Χαρακτηριστικά και Μέθοδοι.....	68
6.1.3. Αποτελέσματα.....	69
6.1.4. Συμπεράσματα.....	71
6.2. Γενικά συμπεράσματα και μελλοντικές κατευθύνσεις.....	72
7. ΕΠΙΛΟΓΟΣ.....	73
8. ΒΙΒΛΙΟΓΡΑΦΙΑ.....	74

ΛΙΣΤΑ ΣΧΗΜΑΤΩΝ

1. Προσεγγίσεις τεχνητής νοημοσύνης.....	16
2. Οι καλύτερες εφικτές βαθμολογίες με κάθε μέθοδο επιλογής.....	24
3. Πραγματικές τιμές εναντίον των τιμών που προβλέπει το μοντέλο γραμμικής παλινδρόμησης.....	29
4. Απεικόνιση της παλινδρόμησης Random Forest.....	36
5. Απεικόνιση των Νευρωνικών Δικτύων.....	36
6. α) NDCG κέρδη εκτός σύνδεσης.....	45
β) Κέρδη στις online κρατήσεις.....	45
7. Αξιολόγηση της απόδοσης του μοντέλου.....	51
8. Επισκόπηση του συστήματος τιμολόγησης.....	58
9. Tokyo.....	59
10. Tahoe, CA.....	60
11. Σύγκριση των πραγματικών τιμών με τις προβλέψεις τιμών για 100 σπίτια.....	65
12. Διαδικασία καθαρισμού, φιλτραρίσματος και τμηματοποίησης των δεδομένων που συλλέγονται από την ιστοσελίδα Inside Airbnb.....	69
13. Κατανομή των συντελεστών που υπολογίζονται με τον αλγόριθμο MLR.....	70
14. Κατανομή των συντελεστών που υπολογίζονται με τον αλγόριθμο SVR.....	71

ΛΙΣΤΑ ΠΙΝΑΚΩΝ

1. Μετρήσεις επιδόσεων των δοκιμασμένων μοντέλων.....	25
2. Συνεχή χαρακτηριστικά με διαφορετικά μοντέλα.....	36
3. Το μοντέλο XGBoost με διαφορετικούς μετασχηματισμούς ετικετών.....	37
4. Αποτελέσματα του μοντέλου XGBoost.....	40
5. Τα 5 πιο σημαντικά χαρακτηριστικά.....	40
6. Τα 5 λιγότερο σημαντικά χαρακτηριστικά.....	40
7. Αποτελέσματα του Νευρωνικού Δικτύου.....	41
8. Αποτελέσματα offline αξιολόγησης.....	54
9. Ακρίβεια των αλγορίθμων MLR και SVR.....	69
10. Συντελεστές που ενεργούν στους 6 δείκτες ικανοποίησης με τη χρήση των αλγορίθμων MLR και SVR.....	70

1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΠΛΑΤΦΟΡΜΑ ΤΗΣ AIRBNB

1.1. Στόχος και περιγραφή της Airbnb

Η Airbnb είναι ιστοσελίδα καταχώρισης, εύρεσης και ενοικίασης καταλυμάτων. Ιδρύθηκε τον Αύγουστο του 2008 από τον Joe Gebbia, τον Brian Chesky και τον Nathan Blecharczyk και η βάση της βρίσκεται στο Σαν Φρανσίσκο της Καλιφόρνια.

Μέσω της Airbnb, ο επισκέπτης μπορεί να βρει μια μεγάλη ποικιλία καταλυμάτων όπως διαμέρισμα, κάστρο, βίλα κ.α., ανάλογη με το μέγεθος της κοινότητας των χρηστών της, οι οποίοι υπόσχονται μοναδικές ταξιδιωτικές εμπειρίες. Το εύρος των τιμών ποικίλλει σημαντικά, όπως και η τοποθεσία του καταλύματος. Από τα στοιχεία που έχουν αναρτηθεί στην ιστοσελίδα, υπάρχουν εγγεγραμμένα καταλύματα σε περισσότερες από 34.000 πόλεις και 191 χώρες (Ιούνιος 2016). Η Airbnb είναι ένας εύκολος και γρήγορος τρόπος να εκμεταλλευθεί οικονομικά ο κάθε χρήστης το χώρο του, παρουσιάζοντάς τον σε ένα κοινό εκατομμυρίων χρηστών.^[1] Δεδομένης της δημοτικότητάς του και της εμβέλειάς του στον τουρισμό και τη βιομηχανία φιλοξενίας, οι ερευνητές έχουν αρχίσει να πραγματοποιούν συστηματικές μελέτες για το φαινόμενο Airbnb. Ένα βασικό χαρακτηριστικό της εμπειρίας χρήστη της Airbnb είναι η ικανότητά του να προσφέρει μια αυθεντική συνάντηση τουριστών-φιλοξενουμένων που δεν μπορεί να αναπαραχθεί σε συμβατικά ξενοδοχεία. Η ταχεία ανάπτυξη της Airbnb, με το ξεχωριστό μοντέλο λειτουργίας της, όχι μόνο προσφέρει μια εναλλακτική εμπειρία διαμονής για τους χρήστες της, αλλά αμφισβητεί επίσης τις θεωρίες και τις πρακτικές που αναπτύχθηκαν από τη συμβατική ξενοδοχειακή βιομηχανία. Η Airbnb αναφέρεται στον εαυτό της ως μια ξεχωριστή διαδικτυακή πλατφόρμα δημιουργώντας «μια αξιόπιστη κοινοτική αγορά για τους χρήστες ώστε να παραθέτουν, να ανακαλύπτουν και να κάνουν κράτηση για μοναδικά καταλύματα σε όλο τον κόσμο» και «συνδέει τους ανθρώπους με μοναδικές ταξιδιωτικές εμπειρίες». Με την υπόσχεση της Airbnb να παρέχει μια μοναδική εμπειρία από άνθρωπο σε άνθρωπο, οι ερευνητές έχουν ξεκινήσει έρευνες για την αναγνώριση των διαστάσεων και χαρακτηριστικών που αποτελούν τη βάση της εμπειρίας της Airbnb.^[2]

1.2. Στατιστικά στοιχεία της Airbnb

Για όλους όσους ψάχνουν μια κατοικία που να θυμίζει το προσωπικό τους σπίτι, ή ακόμα και κάτι καλύτερο, το Airbnb έχει αναδειχθεί ως η βέλτιστη λύση για εύρεση καταλυμάτων ανά τον κόσμο. Από τις αρχές της λειτουργίας της, η συγκεκριμένη πλατφόρμα έχει γνωρίσει τεράστια επιτυχία αφού δραστηριοποιείται σε περισσότερες από 190 χώρες σε ολόκληρο τον κόσμο. Παρόλο τον ανταγωνισμό που υπάρχει από άλλες πλατφόρμες όπως το Booking, το TripAdvisor και το Trivago, η πλατφόρμα της Airbnb έχει καταφέρει να ξεχωρίσει διότι κατέχει πρωτοποριακά χαρακτηριστικά και η χρήση της είναι απλοποιημένη με αποτέλεσμα να έχει

αυξήσει δραματικά τον αριθμό των χρηστών της. Το 97% της τιμής του καταλύματος παραμένει στον ίδιο οικοδεσπότη ενώ η πλατφόρμα δεσμεύει μόλις το 3% για τις υπηρεσίες τις οποίες προσφέρει.^[26]

Η Airbnb κατέγραψε περισσότερες από 40 εκατομμύρια αφίξεις επισκεπτών το 2015. Ήταν μία τεράστια επιτυχία της πλατφόρμας, αφού μόνο σε ένα έτος μπόρεσε και ξεπέρασε το συνολικό αριθμό κρατήσεων από τα προηγούμενα 6 χρόνια.

Το 2016 υπήρξε διπλασιασμός στο προηγούμενο νούμερο, αφού καταγράφηκαν 80 εκατομμύρια αφίξεις επισκεπτών. Το 2017 καταγράφηκαν 115 εκατομμύρια αφίξεις επισκεπτών, το 2018 υπήρξαν 140 εκατομμύρια αφίξεις επισκεπτών και το 2019 υπήρξε ραγδαία αύξηση, αφού καταγράφηκαν 272 εκατομμύρια αφίξεις επισκεπτών. Το 2020 και μετά το χτύπημα του κορονοϊού, οι αφίξεις μειώθηκαν στα 193 εκατομμύρια, γεγονός που έριξε κατά πολύ την αξία της εταιρίας στο χρηματιστήριο, αφού το 2019 έφτασε να κοστολογείται στα 38 δισεκατομμύρια, ενώ το 2020 έπεσε στα 18 δισεκατομμύρια.

Ένα ενδιαφέρον στατιστικό, είναι ότι αυτή την περίοδο υπάρχουν πάνω από ένα εκατομμύριο άνθρωποι οι οποίοι διαμένουν σε ένα κατάλυμα της Airbnb σε οποιαδήποτε δεδομένη νύχτα του έτους. Αυτό ισοδυναμεί με μια κατανεμημένη κινητή πόλη μεσαίου μεγέθους όπως το Σαν Φρανσίσκο ή το Μιλάνο, που κατοικείται από παγκόσμιους πολίτες που θέλουν να εξερευνήσουν τον κόσμο διαμένοντας στο σπίτι κάποιου άλλου ανθρώπου.

Μία μελέτη του 2016 η οποία έγινε από την εταιρία παροχής χρηματοοικονομικών υπηρεσιών Cowen and Company με έδρα τη Νέα Υόρκη εκτιμά ότι οι ετήσιες αφίξεις θα μπορούσαν ενδεχομένως να αυξηθούν και ένα δισεκατομμύριο έως το 2025.^[3]

1.3. Ιστορικά στοιχεία της Airbnb

Το όνομα της Airbnb, προήλθε από το ακρωνύμιο του αρχικού ονόματος Airbedandbreakfast. Η αρχική ιδέα γεννήθηκε, όταν τον Οκτώβριο του 2007 έλαβε χώρα στο Σαν Φρανσίσκο ένα διεθνές συνέδριο σχεδιαστών. Ο Joe Gebbia και ο Brian Chesky, οι οποίοι ήταν τότε συγκατοίκοι, καθώς αδυνατούσαν να πληρώσουν το ενοίκιο στο διαμέρισμα που έμεναν, πήραν την απόφαση να μετατρέψουν το σαλόνι τους σε ξενώνα bed and breakfast, προσφέροντας σε τρεις συμμετέχοντες του συνεδρίου στέγη, φουσκωτά στρώματα (αέρα) και σπιτικό πρωινό έναντι πληρωμής. Μετά τις 3 πρώτες εκμισθώσεις, άρχισαν να λαμβάνουν μηνύματα ηλεκτρονικού ταχυδρομείου για τη διαθεσιμότητα του καταλύματος από μέρη όπως το Μπουένος Άιρες, το Λονδίνο και η Ιαπωνία.

Το καλοκαίρι του 2009, το Airbnb δεν είχε καταφέρει να διεισδύσει στην αγορά της Νέας Υόρκης όσο θα επιθυμούσαν οι δημιουργοί του, και έτσι οι Gebbia και Chesky χρησιμοποίησαν πειραματικά 24 καταλύματα για να καταλάβουν ποιο ήταν το πρόβλημα.

Όπως αποδείχθηκε, οι οικοδεσπότες δεν προσπαθούσαν αρκετά ώστε να αναδείξουν τα καταλύματά τους.

Έτσι λοιπόν, το ζευγάρι των δημιουργών της εταιρίας σκέφτηκε την πιο απλή αλλά αποτελεσματική λύση, νοίκιασε μια επαγγελματική φωτογραφική μηχανή αξίας 5.000 δολλαρίων και πήγε σε σπίτια λαμβάνοντας επαγγελματικές εικόνες από όσο το δυνατόν περισσότερα καταλύματα της Νέας Υόρκης. Οι καινούριες αυτές φωτογραφίες οδήγησαν σε δύο έως τρεις φορές περισσότερες κρατήσεις στις λίστες της Νέας Υόρκης και έως το τέλος του μήνα τα έσοδα του Airbnb στη συγκεκριμένη πόλη είχαν διπλασιαστεί.

Το 2014 αποφασίστηκε από την εταιρία με σκοπό την κερδοφορία της, η επέκτασή της σε παγκόσμια κλίμακα. Ένας από τους τρόπους μέσω των οποίων το Airbnb δοκίμασε να επεκταθεί στην παγκόσμια αγορά, ήταν η δοκιμή στη Γαλλία. Επιλέχθηκε να προωθηθεί τυχαία στις μισές διεθνείς πόλεις με φυσικές διαφημίσεις και στις άλλες μισές μέσω της χρήσης διαφημίσεων από το Facebook. Τελικά, αποδείχθηκε ότι οι διαφημίσεις μέσω Facebook δεν είχαν τα ανάλογα αποτελέσματα και περιορίστηκαν.

Στα τέλη του 2013, αποφασίστηκε να ξεκινήσει εκ νέου η πολιτική referral. Έπειτα από 3 μήνες και 30.000 γραμμές κώδικα, το πρόγραμμα παραπομπών ξεκίνησε ξανά τον Ιανουάριο του 2014. Η εταιρία εξέτασε όλες τις μεταβλητές και αποδείχθηκε ότι όταν η προσφορά από μηνύματα ηλεκτρονικού ταχυδρομείου που έστελνε η Airbnb σε εν δυνάμει άτομα που θα λάμβαναν referrals είχε αλτρουιστικό χαρακτήρα οδηγούσε σε μεγαλύτερο όγκο referrals από τους ίδιους.

Το καινούριο αυτό πρόγραμμα παραπομπών της Airbnb, είχε ως αποτέλεσμα να υπάρξουν εκατοντάδες χιλιάδες διανυκτερεύσεις οι οποίες κρατήθηκαν με παραπομπή των χρηστών το 2014, και σε ορισμένες αγορές οι παραπομπές αυξήθηκαν κατά 25%. Έτσι, διαπιστώθηκε ότι όσοι χρήστες κάνουν χρήση referrals είναι καλύτεροι από το μέσο χρήστη, αφού τείνουν να παραμένουν δεσμευμένοι με την πλατφόρμα και κλείνουν μελλοντικά ταξίδια. Εκτός αυτού, είναι αρκετά πιθανό να στείλουν και οι ίδιοι referrals σε άλλους χρήστες.^[4]

1.4. Λειτουργία της πλατφόρμας

Οι χρήστες της ιστοσελίδας, αρχικά πρέπει να κάνουν εγγραφή, δημιουργώντας ένα προσωπικό διαδικτυακό προφίλ, πριν από τη χρήση της. Κάθε ιδιοκτησία συνδέεται με έναν οικοδεσπότη, το προφίλ του οποίου περιλαμβάνει σχόλια από τους προηγούμενους χρήστες, συστάσεις από τους προηγούμενους θαμώνες, μία αξιολόγηση, καθώς και ένα ιδιωτικό σύστημα μηνυμάτων.

Η λίστα δε θα εμφανίζεται στο δημόσιο ιστό, μέχρι ο ίδιος ο χρήστης να είναι έτοιμος να τη δημοσιοποιήσει. Οι χρήστες μπορούν να καθορίσουν την κοστολόγηση του καταλύματος. Επίσης, οι χρήστες μπορούν, είτε να χρεώνουν διαφορετικές τιμές για κάθε βράδυ, είτε να

θέσουν εβδομαδιαία και μηνιαία διαμονή, είτε ακόμα να καθορίσουν την τιμή σύμφωνα με την εποχή (εποχιακή κοστολόγηση). Τα ένθετα των περιγραφών και των τίτλων, δύναται να χρησιμοποιηθούν με σκοπό τη διαφήμιση του χώρου. Επίσης, οι χρήστες μπορούν να προσδιορίσουν τους κανόνες του σπιτιού, καθώς και να δημοσιοποιήσουν και άλλες περιγραφές σχετικά με την κατοικία. Η Airbnb επιτρέπει στους χρήστες της να δημοσιεύσουν μέχρι και 24 φωτογραφίες του χώρου τους. Στα δικαιώματα που έχει ο οικοδεσπότης, περιλαμβάνεται και μια δωρεάν επαγγελματική φωτογράφιση, η οποία προσφέρεται από την εταιρία. Η όμορφη φωτογραφία είναι ένα από τα κύρια και απαραίτητα χαρακτηριστικά της ιστοσελίδας της Airbnb, καθώς και της εφαρμογής της.^[5] Το Προφίλ, είναι ένα μέρος στο οποίο οι επισκέπτες μπορούν να βρουν περισσότερα στοιχεία για τους οικοδεσπότες, οι οποίοι το χρησιμοποιούν με σκοπό να παρουσιάσουν την ταυτότητά τους δημοσίως, καθώς και για να εκφράσουν τη φιλοσοφία τους σχετικά με την ιδέα της φιλοξενίας.^[6]

Οι επισκέπτες, αρχικά, πρέπει να στείλουν ένα απευθείας μήνυμα στο χρήστη, μέσω της Airbnb, υποβάλλοντας ερωτήσεις σχετικά με το ακίνητο. Οι χρήστες ωστόσο, έχουν τον πλήρη έλεγχο σχετικά με το ποιος κάνει κράτηση του χώρου τους. Όταν ένας πελάτης κάνει ένα αίτημα για κράτηση, ο οικοδεσπότης έχει τουλάχιστον 24 ώρες για να το αποδεχτεί ή να το απορρίψει, σύμφωνα με την κρίση του.^[6]

Τέλος, αφού ο χρήστης αποδεχτεί μία κράτηση, καθίσταται δυνατό να συντονίσει τις ώρες συνάντησης και επικοινωνίας με τους επισκέπτες. Στη συνέχεια και μετά την ολοκλήρωση της κράτησης, οι χρήστες ενθαρρύνονται να αφήσουν την κριτική τους. Οι κριτικές των χρηστών, βοηθούν στην οικοδόμηση της εγκυρότητας και στη δημιουργία αναφορών τόσο για τους επισκέπτες, όσο και για τους οικοδεσπότες.^[6]

1.5. Η διαφορά της μίσθωσης σε σχέση με τα ξενοδοχεία και τις επιχειρήσεις καταλυμάτων

Τα ξενοδοχεία και οι επιχειρήσεις καταλυμάτων, λειτουργούν μέσα σε ένα συγκεκριμένο, καθορισμένο χρονικό πλαίσιο κάθε χρόνο, συχνά ανάλογα με τη γεωγραφική τους θέση. Ενδεικτικά, τα ξενοδοχεία και οι επιχειρήσεις καταλυμάτων που βρίσκονται στην πόλη, λειτουργούν όλο το χρόνο, ενώ τα ξενοδοχεία και οι επιχειρήσεις καταλυμάτων σε νησιωτικούς προορισμούς, εξαιτίας της εποχικότητας του τουρισμού, λειτουργούν συνήθως ορισμένους μήνες. Σε κάθε περίπτωση όμως υπάρχει ένα σαφές χρονικό πλαίσιο λειτουργίας, το οποίο είναι καθορισμένο. Ενδεχομένως να υπάρχει κάποια παράταση ή σμίκρυνσή του για τα εποχιακά ξενοδοχεία, αλλά οι αποκλίσεις από τις αρχικές προαναγγελθείσες ημέρες λειτουργίας είναι μικρές. Επιπλέον, δεν προβλέπεται η μη λειτουργία μέσα στη σεζόν, δηλαδή το ξενοδοχείο ή η επιχείρηση καταλυμάτων θα λειτουργεί όλες τις ημέρες της σεζόν.

Αντιθέτως, η προσφορά καταλυμάτων μέσω της πλατφόρμας Airbnb, είναι διαρκώς μεταβαλλόμενη, αναλόγως από τη ζήτηση. Για παράδειγμα, ενδέχεται κάποιο κατάλυμα να εμφανίζεται διαθέσιμο προς μίσθωση στις πλατφόρμες για κάποια συγκεκριμένη χρονική περίοδο ή ακόμα και για ελάχιστες ημέρες, αλλά μετά να αποσύρεται από την πλατφόρμα. Αυτό κυρίως συμβαίνει, διότι υπάρχει προσαρμογή της προσφοράς στη ζήτηση, δηλαδή σε περιόδους όπου υπάρχει μεγάλη ζήτηση (είτε εποχικά είτε για κάποιο σημαντικό γεγονός) εμφανίζονται περισσότερα καταλύματα προς μίσθωση στις πλατφόρμες, ενώ σε περιόδους χαμηλής ζήτησης, πολλά από αυτά αποσύρονται από την πλατφόρμα. Επιπλέον, ο εκάστοτε ιδιοκτήτης, υπάρχει περίπτωση να κατοικεί ο ίδιος στο εν λόγω κατάλυμα και να το διαθέτει για εκμίσθωση σε κάποιον τρίτο τις ημέρες τις οποίες ο ίδιος έχει προγραμματίσει να απουσιάζει, αποκομίζοντας με αυτό τον τρόπο ένα επιπλέον εισόδημα, χωρίς να ενδιαφέρεται για περαιτέρω μίσθωση.

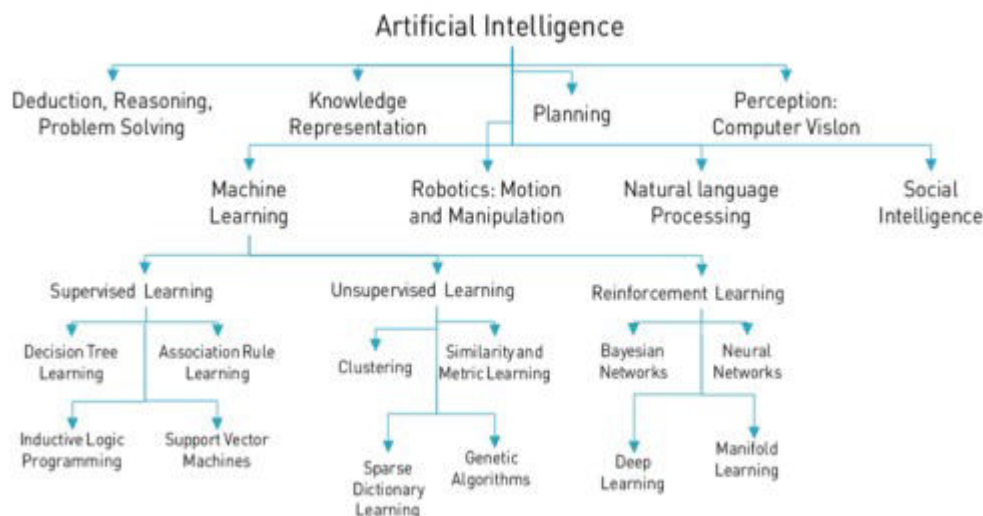
Ιδιαίτερο ενδιαφέρον παρουσιάζει το πόσο συνυφασμένη είναι η πληρότητα και οι δείκτες εσόδων με τον εκάστοτε προορισμό. Για παράδειγμα, στην Αθήνα υπάρχει υψηλή ζήτηση, αλλά λόγω αυξημένης προσφοράς καταλυμάτων, ο ανταγωνισμός είναι μεγάλος, οδηγώντας σε χαμηλές προσφερόμενες τιμές. Αντίθετα, σε νησιά όπως η Μύκονος και η Σαντορίνη, όπου τα διαθέσιμα καταλύματα είναι εκ των πραγμάτων πεπερασμένα εξαιτίας του οικοδομικού κορεσμού και το τουριστικό “brand” τους ιδιαίτερα υψηλό, καταγράφεται υπερβολικά μεγάλη πληρότητα. Αυτό οφείλεται στο γεγονός ότι η ζήτηση υπάρχει ήδη, ενώ καθίσταται αδύνατο να υπάρξουν περισσότερα καταλύματα. Ως επακόλουθο της εν λόγω εξέλιξης, οι τιμές είναι ιδιαίτερα αυξημένες, καθώς ο ανταγωνισμός περνά στους υποψήφιους μισθωτές.^[7]

2. ΕΙΣΑΓΩΓΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

2.1. Εισαγωγή στην Τεχνητή Νοημοσύνη

Η τεχνητή νοημοσύνη ορίζεται ως ένας τομέας μελέτης στο πεδίο της επιστήμης των υπολογιστών. Αφορά την ανάπτυξη υπολογιστικών συστημάτων τα οποία είναι ικανά να συμπεριφέρονται όπως ο ανθρώπινος εγκέφαλος, συμμετέχοντας σε διαδικασίες μάθησης, συλλογισμού και αυτο-διόρθωσης. Ένας από τους βασικούς θεμελιωτές της τεχνητής νοημοσύνης, ο John McCarthy, περιέγραψε την τεχνολογία ως μια επιστήμη και τεχνική κατασκευής ευφύων μηχανών και ειδικά έξυπνων προγραμμάτων υπολογιστών. Η τεχνητή νοημοσύνη, έχει παρόμοια λειτουργία με την ανθρώπινη νοημοσύνη, που θέλει να μαθαίνει, να προσαρμόζεται, να αναγνωρίζει και να διορθώνει, και να επεξεργάζεται πληροφορίες ώστε να παρέχει αποτελέσματα με τον ίδιο τρόπο όπως το δυναμικό ενός οργανισμού. Η τεχνητή νοημοσύνη ταξινομείται σε τέσσερις μεγάλες προσεγγίσεις:

- **Μηχανισμός συστήματος που δρα σαν άνθρωπος**, δηλαδή η τέχνη της δημιουργίας μηχανών οι οποίες εκτελούν λειτουργίες που απαιτούν νοημοσύνη όταν εκτελούνται από ανθρώπους.
- **Μηχανισμός συστήματος που σκέφτεται σαν άνθρωπος**, δηλαδή η κατασκευή ευφύων μηχανών μέσω της προσέγγισης γνωστικής μοντελοποίησης, όπου γίνονται δοκιμές οι οποίες είναι προγραμματισμένες με ένα συγκεκριμένο τρόπο ανθρώπινης σκέψης, με σκοπό να μετατραπεί αυτή η σκέψη σε μηχανική.
- **Μηχανισμός συστήματος που σκέφτεται λογικά**, δηλαδή η μελέτη των υπολογισμών που καθιστούν δυνατή την αντίληψη, τη λογική και τη δράση. Για τον έλεγχο της λογικής σκέψης των μηχανών χρησιμοποιούνται οι νόμοι της προσέγγισης σκέψης.
- **Μηχανισμός συστήματος που δρα λογικά**, δηλαδή η αυτοματοποίηση της συμπεριφοράς των ευφύων μηχανών με τέτοιο τρόπο, ώστε οι συγκεκριμένες μηχανές να έχουν αντίληψη και να ενεργούν. Έχει μεγάλη σημασία για την επίλυση πραγματικών προβλημάτων.^[8]



Σχήμα 1. Προσεγγίσεις τεχνητής νοημοσύνης.^[9]

2.2. Ορισμός της Μηχανικής Μάθησης

Μηχανική μάθηση / Machine learning, είναι υποπεδίο της επιστήμης των υπολογιστών το οποίο αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. Διερευνά τη μελέτη και την κατασκευή αλγορίθμων οι οποίοι μπορούν να μαθαίνουν από τα δεδομένα τους και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν δημιουργώντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις οι οποίες βασίζονται στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα.

Η μηχανική μάθηση είναι στενά συνδεδεμένη και συχνά συγχέεται με την υπολογιστική στατιστική, ένας κλάδος ο οποίος επίσης επικεντρώνεται στην πρόβλεψη μέσω της χρήσης υπολογιστών. Έχει ισχυρούς δεσμούς με τη μαθηματική βελτιστοποίηση, η οποία της παρέχει μεθόδους, θεωρία και τομείς εφαρμογής.

Εφαρμόζεται σε μία σειρά από υπολογιστικές εργασίες, στις οποίες τόσο ο σχεδιασμός όσο και ο ρητός προγραμματισμός των αλγορίθμων είναι αδύνατος. Παραδείγματα εφαρμογών αποτελούν τα φίλτρα spam (spam filtering), η οπτική αναγνώριση χαρακτήρων (OCR), οι μηχανές αναζήτησης και η υπολογιστική όραση. Μερικές φορές, η μηχανική μάθηση συγχέεται με την εξόρυξη δεδομένων, όπου η τελευταία επικεντρώνεται περισσότερο στην εξερευνητική ανάλυση των δεδομένων, που είναι γνωστή και ως μη επιτηρούμενη μάθηση.^[10]

2.3. Υπολογιστικές μηχανές και νοημοσύνη

Ο Tom M. Mitchell πρότεινε έναν πιο επίσημο ορισμό που χρησιμοποιείται παντού : «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P , αν η επίδοσή του σε εργασίες της κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία E ».^[11]

Ο παραπάνω ορισμός είναι σημαντικός για τον καθορισμό της μηχανικής μάθησης σε βασικό λειτουργικό πλαίσιο παρά με γνωστικούς όρους, ακολουθώντας έτσι την πρόταση του Alan Turing στην εργασία του «υπολογιστικές μηχανές και Νοημοσύνη», ότι το ερώτημα αν μπορούν οι μηχανές να σκεφτούν, μπορεί να αντικατασταθεί με το ερώτημα αν μπορούν οι μηχανές να κάνουν αυτό που εμείς (ως οντότητες που σκέφτονται) μπορούμε να κάνουμε.^[12]

Στο αναλυτικό πεδίο δεδομένων, η μηχανική μάθηση είναι μια μέθοδος η οποία χρησιμοποιείται για την επινόηση πολύπλοκων μοντέλων και αλγορίθμων που κατευθύνουν στην πρόβλεψη. Τα αναλυτικά μοντέλα επιτρέπουν στους ερευνητές, τους επιστήμονες δεδομένων (data scientists), τους μηχανικούς και τους αναλυτές να παράγουν αξιόπιστες αποφάσεις και αποτελέσματα και να αναδεικνύουν αλληλοσυσχετίσεις μέσω της μάθησης από ιστορικές σχέσεις και τάσεις στα υπάρχοντα δεδομένα.

2.4. Τύποι προβλημάτων και εργασιών

Συνήθως, οι εργασίες στη μηχανική μάθηση ταξινομούνται σε τρεις μεγάλες κατηγορίες, ανάλογα με τη φύση του εκπαιδευτικού «σήματος» ή την «ανατροφοδότηση» που είναι διαθέσιμα σε ένα σύστημα εκμάθησης. Αυτές είναι οι παρακάτω:

- **Επιτηρούμενη μάθηση** (supervised learning): Το υπολογιστικό πρόγραμμα δέχεται τις παραδειγματικές εισόδους καθώς και τα επιθυμητά αποτελέσματα από έναν «δάσκαλο», και ο στόχος είναι να μάθει ένα γενικό κανόνα έτσι ώστε να ταιριάζει τις εισόδους μαζί με τα αποτελέσματα.
- **Μη επιτηρούμενη μάθηση** (unsupervised learning): Ο αλγόριθμος μάθησης πρέπει να βρει τη δομή των δεδομένων εισόδου, χωρίς να του παρέχεται κάποια εμπειρία. Η μη επιτηρούμενη μάθηση μπορεί να είναι αυτοσκοπός (ανακαλύπτοντας κρυμμένα μοτίβα σε διάφορα δεδομένα) ή μέσο για ένα τέλος.
- **Ενισχυτική μάθηση**: Ένα υπολογιστικό πρόγραμμα και ένα δυναμικό περιβάλλον αλληλεπιδρούν μεταξύ τους έτσι ώστε να επιτευχθεί ένας συγκεκριμένος στόχος, χωρίς κάποιος δάσκαλος να του λέει ρητά αν έχει πλησιάσει σε αυτόν τον στόχο. Ένα άλλο παράδειγμα είναι να μάθει να παίζει ένα παιχνίδι εναντίον κάποιου αντιπάλου.
- Ανάμεσα στην επιτηρούμενη και στη μη επιτηρούμενη μάθηση βρίσκεται η **ημι-επιτηρούμενη μάθηση**, στην οποία ο δάσκαλος δίνει ένα ελλιπές εκπαιδευτικό σήμα,

δηλαδή ένα σύνολο εκπαίδευσης με κάποια από τα αποτελέσματα να λείπουν. Μία ειδική περίπτωση της αρχής αυτής είναι η **Μεταγωγή**, όπου το σύνολο των καταστάσεων του προβλήματος είναι γνωστό κατά το χρόνο εκμάθησης, όμως λείπει ένα μέρος των στόχων.

- Μεταξύ άλλων κατηγοριών μηχανικής μάθησης, υπάρχει και η **διαδικασία εκμάθησης (meta learning)**, η οποία μαθαίνει στη μηχανή τις δικές της επαγωγικές μεθόδους, βασιζόμενη στην προηγούμενη εμπειρία.
- Η **Αναπτυξιακή μάθηση** (Developmental robotics), έχει αναπτυχθεί για την εκμάθηση από ρομπότ και δημιουργεί τη δική της ακολουθία μαθησιακών καταστάσεων, ώστε το ρομπότ συσσωρευτικά αποκτάει ποικιλία δεξιοτήτων μέσω της αυτόνομης αυτοεξερεύνησης και της κοινωνικής αλληλεπίδρασης με ανθρώπους εκπαιδευτές και χρησιμοποιεί μηχανισμούς καθοδήγησης, όπως η ενεργητική μάθηση, η ωρίμανση και η μίμηση.^[13]

Ακόμη μία κατηγοριοποίηση των προβλημάτων μηχανικής μάθησης προκύπτει όταν κάποιος θεωρήσει το επιθυμητό αποτέλεσμα του συστήματος μηχανικής μάθησης:

- Στην **ταξινόμηση**, τα δεδομένα εισόδου χωρίζονται σε δύο ή σε περισσότερες κλάσεις και η μηχανή πρέπει να κατασκευάσει ένα μοντέλο, το οποίο θα αντιστοιχίζει τα δεδομένα σε μία ή περισσότερες κλάσεις. Συνήθως αυτό εμπίπτει στην επιτηρούμενη μάθηση. Τα φίλτρα spam είναι ένα παράδειγμα ταξινόμησης, στα οποία οι εισοδοί είναι τα emails ή άλλα μηνύματα και οι κλάσεις χωρίζονται σε “spam” και “όχι spam”.
- Στην **παλινδρόμηση**, η οποία είναι επίσης ένα πρόβλημα επιτηρούμενης μάθησης, τα αποτελέσματα είναι συνεχή και όχι διακριτά.
- Στη **συσταδοποίηση**, ένα σύνολο εισόδων πρόκειται να χωριστεί σε ομάδες. Αντιθέτως με την ταξινόμηση, οι ομάδες δεν είναι γνωστές εκ των προτέρων, καθιστώντας το συγκεκριμένο διαχωρισμό μία τυπική εργασία μη επιτηρούμενης μάθησης.
- Στην **εκτίμηση πυκνότητας** βρίσκει την κατανομή των δεδομένων εισόδου σε κάποιο χώρο.
- Σε προβλήματα **μείωσης διαστασιμότητας** (dimensionality reduction), τα όποια δεδομένα απλοποιούνται και αντιστοιχίζονται σε ένα χώρο με λιγότερες διαστάσεις. Το στατικό μοντέλο θεμάτων είναι ένα σχετικό πρόβλημα, στο οποίο η μηχανική μάθηση καλείται να βρει έγγραφα που καλύπτουν παρόμοια θέματα από ένα σύνολο εγγράφων που είναι γραμμένα σε φυσική γλώσσα.^[13]

2.5. Προσεγγίσεις

2.5.1. Εκμάθηση με δέντρο απόφασης

Η εκμάθηση με δέντρο απόφασης χρησιμοποιεί ένα δέντρο απόφασης ως προγνωστικό μοντέλο, το οποίο αντιστοιχίζει παρατηρήσεις σχετικά με ένα στοιχείο σε συμπεράσματα τα οποία είναι σχετικά με την τιμή στόχο του αντικειμένου.

2.5.2. Εκμάθηση με κανόνες συσχέτισης

Η εκμάθηση με κανόνες συσχέτισης είναι μία μέθοδος ανακάλυψης σχέσεων οι οποίες είναι ενδιαφέρουσες μεταξύ των μεταβλητών σε μεγάλες βάσεις δεδομένων.

2.5.3. Τεχνητά νευρωνικά δίκτυα

Ένας αλγόριθμος εκμάθησης τεχνητού νευρωνικού δικτύου, ο οποίος ονομάζεται συνήθως “νευρωνικό δίκτυο” (NN), είναι ένας αλγόριθμος μάθησης, που εμπνέεται από τις λειτουργικές πτυχές και τη δομή των βιολογικών νευρωνικών δικτύων.

Η δομή των υπολογισμών βασίζεται σε μια ομάδα εσωτερικά διασυνδεδεμένων τεχνητών νευρώνων, οι οποίοι επεξεργάζονται την πληροφορία και έπειτα κάνουν υπολογισμούς επικοινωνώντας μεταξύ τους. Τα σύγχρονα νευρωνικά δίκτυα είναι εργαλεία μη γραμμικής στατιστικής μοντελοποίησης δεδομένων.

Συνήθως χρησιμοποιούνται για τη μοντελοποίηση πολύπλοκων σχέσεων μεταξύ δεδομένων εισόδου και εξόδου, για την ανακάλυψη προτύπων στα δεδομένα αυτά, ή για τον εντοπισμό στατιστικής δομής σε μία άγνωστη κοινή κατανομή πιθανότητας μεταξύ των μεταβλητών που παρατηρούνται.

2.5.4. Βαθιά μάθηση / Deep learning

Η ανάπτυξη των καρτών γραφικών για προσωπική χρήση, καθώς και η πτώση των τιμών του υλικού που παρατηρήθηκε τα τελευταία χρόνια, οδήγησε στην ανάπτυξη της ιδέας της Βαθιάς μάθησης. Η συγκεκριμένη προσέγγιση προσπαθεί να μοντελοποιήσει τον τρόπο που ο ανθρώπινος εγκέφαλος επεξεργάζεται το φως και τον ήχο και τα μετατρέπει σε όραση και ακοή. Μερικές επιτυχείς εφαρμογές της Βαθιάς μάθησης είναι η μηχανική όραση και η αναγνώριση ομιλίας.

2.5.5. Επαγωγικός λογικός προγραμματισμός

Ο Επαγωγικός λογικός προγραμματισμός (ILP) είναι μια προσέγγιση που διέπει τη μάθηση και χρησιμοποιεί λογικό προγραμματισμό ως τρόπο παρουσίασης των παραδειγμάτων εισόδου, του γνωστικού υποβάθρου και των υποθέσεων.

Δεδομένης μιας κωδικοποίησης του γνωστικού υποβάθρου και ενός συνόλου παραδειγμάτων τα οποία παρουσιάζονται σαν λογική βάση γεγονότων, το υποτιθέμενο λογικό πρόγραμμα που περιέχει όλα τα θετικά και κανένα αρνητικό παράδειγμα, παράγεται από το σύστημα ΕΛΠ.

Ο επαγωγικός προγραμματισμός είναι ένας σχετικός τομέας στον οποίο λαμβάνεται υπόψιν κάθε είδος γλώσσας προγραμματισμού για την αναπαράσταση υποθέσεων (και όχι μόνο λογικό προγραμματισμό), όπως τα συναρτησιακά προγράμματα.

2.5.6. Μηχανές διανυσμάτων υποστήριξης

Οι μηχανές διανυσμάτων υποστήριξης είναι ένα σύνολο μεθόδων επιτηρούμενης μάθησης που χρησιμοποιούνται για την ταξινόμηση και την παλινδρόμηση. Στη συγκεκριμένη περίπτωση δίνεται ένα σύνολο παραδειγμάτων εκπαίδευσης και κάθε φορά δηλώνεται σε ποια από τις δύο κατηγορίες ανήκει το συγκεκριμένο παράδειγμα.

Μία μηχανή διανυσμάτων υποστήριξης κατασκευάζει ένα μοντέλο το οποίο προβλέπει αν το καινούριο παράδειγμα ταιριάζει στη μία κατηγορία ή στην άλλη.

2.5.7. Ομαδοποίηση

Ως ομαδοποίηση, ονομάζεται η διαδικασία κατά την οποία ένα σύνολο παρατηρήσεων χωρίζεται σε υποσύνολα έτσι ώστε αυτές οι παρατηρήσεις που ανήκουν στην ίδια ομάδα (cluster) είναι όμοιες, σύμφωνα με κάποιο ή κάποια προκαθορισμένα κριτήρια, ενώ αντιθέτως οι παρατηρήσεις που προέρχονται από διαφορετικά υποσύνολα είναι ανόμοιες.

Διαφορετικές τεχνικές κατηγοριοποίησης οδηγούν σε διαφορετικές υποθέσεις σχετικά με τη δομή των δεδομένων, οι οποίες συχνά καθορίζονται από κάποιο μέτρο ομοιότητας και αξιολογούνται για παράδειγμα ως προς την εσωτερική συνοχή (ομοιότητα μεταξύ των μελών του ίδιου cluster) και το διαχωρισμό ανάμεσα σε διαφορετικές ομάδες.

Άλλες μέθοδοι που υπάρχουν, βασίζονται στην εκτιμώμενη πυκνότητα και τη συνεκτικότητα των γραφημάτων. Η ομαδοποίηση είναι μία μέθοδος μη επιτηρούμενης μάθησης και μία τεχνική η οποία χρησιμοποιείται επίσης στη στατιστική ανάλυση δεδομένων.

2.5.8. Δίκτυα Bayes

Ένα δίκτυο Bayes, ένα δίκτυο εμπιστοσύνης ή ένα άκυκλο γραφικό μοντέλο είναι ένα πιθανοθεωρητικό γραφικό μοντέλο που απεικονίζει ένα σύνολο τυχαίων μεταβλητών και τη μεταξύ τους υποθετική ανεξαρτησία διαμέσου ενός κατευθυνόμενου άκυκλου γράφου.

Για παράδειγμα, ένα δίκτυο Bayes καθίσταται δυνατό να αναπαραστήσει την πιθανοθεωρητική σχέση μεταξύ ασθενειών και συμπτωμάτων. Δεδομένων των συμπτωμάτων, το δίκτυο μπορεί να χρησιμοποιηθεί με στόχο τον υπολογισμό των πιθανοτήτων παρουσίας διαφόρων ασθενειών.

2.5.9. Ενισχυτική μάθηση

Η γενικότερη ασχολία της ενισχυτικής μάθησης έχει να κάνει με το πώς ένα υποκείμενο (πράκτορας) θα πρέπει να δράσει σε ένα περιβάλλον, έτσι ώστε να μεγιστοποιηθεί κάποια έννοια μακροπρόθεσμης ανταμοιβής.

Οι αλγόριθμοι ενισχυτικής μάθησης προσπαθούν να βρουν μια πολιτική που αντιστοιχίζει τις καταστάσεις του περιβάλλοντος με τις ενέργειες που ο πράκτορας θα πρέπει να επιτελέσει σε αυτές τις καταστάσεις.

Η ενισχυτική μάθηση διαφέρει από τα προβλήματα επιτηρούμενης μάθησης διότι τα σωστά ζεύγη δεδομένων εισόδου/εξόδου δεν παρουσιάστηκαν ποτέ, αλλά ούτε και οι βέλτιστες δυνατές ενέργειες διορθώθηκαν ρητά.

2.5.10. Εκμάθηση με μέτρο ομοιότητας

Στη συγκεκριμένη κατηγορία προβλημάτων δίνονται στη μηχανή μάθησης ζεύγη παραδειγμάτων τα οποία θεωρούνται όμοια και ζεύγη που θεωρούνται ανόμοια. Έπειτα, η μηχανή μάθησης πρέπει να μάθει μια συνάρτηση ομοιότητας (ή μια συνάρτηση μετρικής απόστασης), που μπορεί να προβλέψει αν δύο καινούρια αντικείμενα είναι όμοια. Πρόκειται για μια τεχνική η οποία γίνεται χρήσιμη σε συστήματα σύστασης.

2.5.11. Γενετικοί αλγόριθμοι

Ένας γενετικός αλγόριθμος (GA) είναι μια ευρετική αναζήτηση η οποία μιμείται τη διαδικασία της φυσικής επιλογής, και χρησιμοποιεί μεθόδους σαν αυτή της μετάλλαξης και της διασταύρωσης έτσι ώστε να δημιουργήσει καινούρια γονότυπα με την ελπίδα εύρεσης αποτελεσματικών λύσεων σε ένα συγκεκριμένο πρόβλημα. Στη μηχανική μάθηση, χρησιμοποιήθηκαν γενετικοί αλγόριθμοι τη δεκαετία του 1980 και του 1990. Αντίστροφα,

τεχνικές μηχανικής μάθησης έχουν χρησιμοποιηθεί για τη βελτίωση της απόδοσης γενετικών και εξελικτικών αλγορίθμων.^[10]

3. ΠΡΟΒΛΕΨΗ ΤΙΜΩΝ

3.1. Πρόβλεψη τιμών στην Airbnb με χρήση μηχανικής μάθησης και ανάλυσης συναισθημάτων.^[14]

3.1.1. Εισαγωγή

Στη συγκεκριμένη έρευνα, η οποία πραγματοποιήθηκε το 2019 και έχει να κάνει με την πρόβλεψη τιμών στην Airbnb με χρήση μηχανικής μάθησης και ανάλυσης συναισθημάτων, από τους Pouya Rezazadeh Kalehbasti, Liubov Nikolenko και Hoormazd Rezaei, αναλύεται η τιμολόγηση ενός ενοικιαζόμενου ακινήτου στην Airbnb η οποία καθίσταται μία δύσκολη εργασία για τον ιδιοκτήτη, καθώς καθορίζει τον αριθμό των πελατών για το συγκεκριμένο μέρος. Από την άλλη, οι πελάτες πρέπει να αξιολογήσουν μία προσφερόμενη τιμή από τον ιδιοκτήτη, έχοντας ελάχιστη γνώση σχετικά με την αξία του ακινήτου. Σκοπός της μηχανικής μάθησης είναι να αναπτυχθεί ένα μοντέλο πρόβλεψης τιμών, χρησιμοποιώντας τεχνικές μηχανικής μάθησης, βαθιά μάθηση και επεξεργασία της φυσικής γλώσσας, τόσο στους ιδιοκτήτες των ακινήτων όσο και στους πελάτες, δίνοντάς τους ελάχιστα δεδομένα σχετικά με το ακίνητο. Οι αξιολογήσεις περιλαμβάνουν χαρακτηριστικά των ενοικίων, των ιδιοκτητών και των πελατών, μέσα από προγνωστικούς παράγοντες και διάφορες μεθόδους, από γραμμική παλινδρόμηση έως μοντέλα βασισμένα σε δέντρα, SVR, ομαδοποίηση K-μέσων (KMC) και νευρωνικά δίκτυα (NN), τα οποία χρησιμοποιούνται για τη δημιουργία ενός μοντέλου πρόβλεψης.

3.1.2. Χαρακτηριστικά και Μέθοδοι

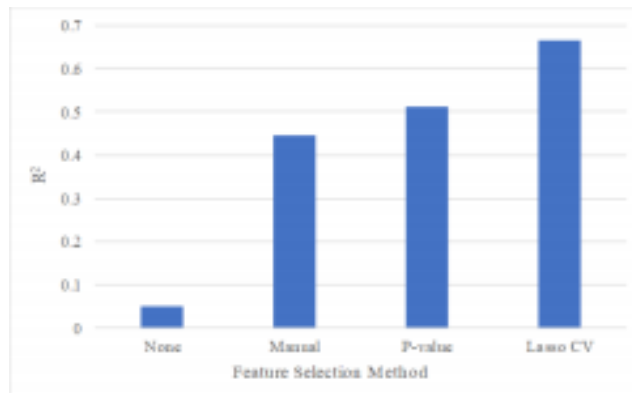
Οι Pouya Rezazadeh Kalehbasti, Liubov Nikolenko και Hoormazd Rezaei, στην έρευνά τους πάνω στην πρόβλεψη τιμών στην Airbnb με χρήση μηχανικής μάθησης και ανάλυσης συναισθημάτων, μετά την προεπεξεργασία των δεδομένων που κάνανε, κατέληξαν ότι το διάνυσμα δυνατοτήτων θα περιέχει 764 στοιχεία. Η τροφοδοσία αυτού του υπερβολικού συνόλου χαρακτηριστικών στα μοντέλα, είχε ως αποτέλεσμα μεγάλη διακύμανση σφάλματος. Κατά συνέπεια, χρησιμοποιήθηκαν πολλές τεχνικές επιλογής χαρακτηριστικών για την εύρεση των δυνατοτήτων με τις πιο προβλεπόμενες τιμές, τόσο για τη μείωση των μεταβλητών του μοντέλου, όσο και για τη μείωση του χρόνου υπολογισμού. Με βάση την προηγούμενη εμπειρία των ερευνητών σχετικά με την εκτίμηση τιμών ακινήτων, η πρώτη μέθοδος που δοκιμάστηκε ήταν η χειροκίνητη επιλογή χαρακτηριστικών για τη δημιουργία μιας γραμμής βάσης για την αξιολόγηση των άλλων διαδικασιών επιλογής χαρακτηριστικών.

Η δεύτερη μέθοδος επιλογής ήταν ο συντονισμός του συντελεστή γραμμικού μοντέλου παλινδρόμησης με Lasso Regularization. Με βάση αυτή την ανάλυση, επιλέχθηκε το μοντέλο με τη βέλτιστη απόδοση έναντι του διαχωρισμού επικύρωσης. Το σύνολο που προέκυψε

αποτελείται από 78 δυνατότητες με μη μηδενικές τιμές, δηλαδή 90% μικρότερο από τον αριθμό των αρχικών χαρακτηριστικών.

Τέλος, χρησιμοποιήθηκαν οι χαμηλότερες τιμές p του μοντέλου τακτικής γραμμικής παλινδρόμησης (Regular Linear Regression Model) για την επιλογή του τρίτου συνόλου χαρακτηριστικών. Στη διαδικασία επιλογής επιβλήθηκε ανώτατο όριο 100 χαρακτηριστικών. Το τελικό σετ αποτελούνταν από 22 χαρακτηριστικά για τα οποία το μοντέλο γραμμικής παλινδρόμησης είχε την καλύτερη απόδοση στο διαχωρισμό επικύρωσης.

Η απόδοση των χειροκίνητα επιλεγμένων χαρακτηριστικών, καθώς και η τιμή p και τα σχήματα επιλογής δυνατοτήτων Lasso συγκρίθηκαν χρησιμοποιώντας τη βαθμολογία R^2 των μοντέλων γραμμικής παλινδρόμησης που έχουν εκπαιδευτεί στο σετ επικύρωσης. Όλα τα μοντέλα υπερτερούν του βασικού μοντέλου, το οποίο χρησιμοποίησε ολόκληρο το σύνολο χαρακτηριστικών και η δεύτερη μέθοδος, η κανονικοποίηση του Lasso, απέδωσε την υψηλότερη βαθμολογία R^2 . Το Σχήμα 1 δείχνει τις καλύτερες βαθμολογίες R^2 που λαμβάνονται χρησιμοποιώντας το σύνολο των χαρακτηριστικών που προσδιορίζονται με κάθε μέθοδο επιλογής δυνατοτήτων.



Σχήμα 2. Οι καλύτερες εφικτές βαθμολογίες με κάθε μέθοδο επιλογής.^[14]

Η γραμμική παλινδρόμηση χρησιμοποιώντας ολόκληρο το σύνολο των χαρακτηριστικών, ελήφθη ως το βασικό μοντέλο για την αξιολόγηση της απόδοσης των άλλων μεθόδων. Μετά την επιλογή ενός συνόλου χαρακτηριστικών με τη χρήση Lasso, εξετάστηκαν διάφορα μοντέλα μηχανικής μάθησης προκειμένου να βρεθεί το βέλτιστο. Όλα τα μοντέλα εκτός από τα νευρωνικά δίκτυα εφαρμόστηκαν χρησιμοποιώντας τη βιβλιοθήκη Scikit-learn. Το μοντέλο νευρωνικού δικτύου υλοποιήθηκε με τη βοήθεια της βιβλιοθήκης Keras. Οι μέθοδοι που χρησιμοποιήθηκαν στη συγκεκριμένη έρευνα είναι οι Ridge Regression, K-means Clustering με Ridge Regression, Support Vector Regression (SVR), Neural Network και Gradient Boosting Tree Ensemble.

3.1.3. Αποτελέσματα

Στη συγκεκριμένη έρευνα, οι τρεις ερευνητές χρησιμοποίησαν το Mean Absolute Error (MAE), το Mean Square Error (MSE) και το R^2 με σκοπό να αξιολογήσουν τα μοντέλα που δοκιμάστηκαν. Στον Πίνακα 1 φαίνονται τα αποτελέσματα των προαναφερθέντων μοντέλων.

Model Name	train split			test split		
	MAE	MSE	R^2 Score	MAE	MSE	R^2 Score
Linear Reg. (Baseline)	0.2744	0.1480	0.690	96895.82	2.4E13	-5.1E13
Ridge Reg.	0.2813	0.15461	0.6765	0.2936	0.1613	0.6601
Gradient Boost	0.2492	0.1376	0.7121	0.3282	0.1963	0.5864
K-means + Ridge Reg.	0.2717	0.1438	0.6992	0.2850	0.1543	0.6748
SVR	0.2132	0.1067	0.7768	0.2761	0.1471	0.6901
Neural Net	0.2602	0.1316	0.7246	0.2881	0.1570	0.6692

Πίνακας 1. Μετρήσεις επιδόσεων των δοκιμασμένων μοντέλων.^[14]

3.1.4. Συμπεράσματα

Οι επιστήμονες που κάνανε τη συγκεκριμένη έρευνα, προσπάθησαν να βρουν το μοντέλο με τις καλύτερες επιδόσεις για την πρόβλεψη τιμών της *Airbnb* με βάση ένα περιορισμένο σύνολο λειτουργιών, συμπεριλαμβανομένων των προδιαγραφών καταλύματος, των πληροφοριών του κατόχου και των κριτικών των πελατών σχετικά με τις καταχωρίσεις. Τεχνικές μηχανικής μάθησης, όπως γραμμική παλινδρόμηση, *tree-based* μοντέλα, SVR και νευρωνικά δίκτυα, μαζί με αναλύσεις των χαρακτηριστικών χρησιμοποιήθηκαν για την επίτευξη των καλύτερων αποτελεσμάτων όσον αφορά το Mean Squared Error (MSE), το Mean Absolute Error (MAE) και το R^2 . Η ανάλυση σπουδαιότητας των χαρακτηριστικών που βασίζεται σε Lasso, μείωσε τη διακύμανση και η χρήση προηγμένων μοντέλων όπως το SVR και τα νευρωνικά δίκτυα είχε ως αποτέλεσμα υψηλότερη βαθμολογία R^2 , τόσο για τα σύνολα επικύρωσης, όσο και για τα σύνολα δοκιμών. Μεταξύ των μοντέλων που δοκιμάστηκαν, το Support Vector Regression (SVR) απέδωσε καλύτερα και απέδωσε βαθμολογία R^2 69% και MSE 0,147 στο σετ δοκιμών. Το συγκεκριμένο επίπεδο ακρίβειας είναι ένα ελπιδοφόρο αποτέλεσμα, δεδομένης της ανομοιογένειας του συνόλου δεδομένων και των σχετικών κρυφών παραγόντων, όπως τα προσωπικά χαρακτηριστικά των ιδιοκτητών, τα οποία ήταν αδύνατο να ληφθούν υπόψη.

3.2. Ανάλυση των τιμών στην Airbnb χρησιμοποιώντας τεχνικές μηχανικής μάθησης.^[15]

3.2.1. Εισαγωγή

Στην παρούσα έρευνα, η οποία έχει σχέση με την ανάλυση των τιμών στην Airbnb χρησιμοποιώντας τεχνικές μηχανικής μάθησης και έγινε στην Καλιφόρνια από τους ερευνητές Jasleen Dhillon, Nandana Priyanka Eluri, Damanpreet Kaur, Aafreen Chhipa, Ashwin Gadupudi, Rajeswari Cherupulli Eravi και Matin Pirouz, πραγματοποιήθηκε αρχικά καθαρισμός και προεπεξεργασία δεδομένων, ενώ στη συνέχεια, πραγματοποιείται περιγραφική, περιοριστική και διερευνητική ανάλυση για να κατανοηθεί καλύτερα η φύση των δεδομένων. Αυτές οι αναλύσεις βοηθάνε ώστε να κατανοηθούν τα σημαντικά χαρακτηριστικά που πρέπει να ληφθούν υπόψη για την πρόβλεψη των τιμών των καταχωρίσεων στην Airbnb. Ακόμη και μετά την εκτέλεση καθαρισμού των δεδομένων στο σύνολο των δεδομένων, μπορεί να υπάρχουν κάποιες ακραίες τιμές που πρέπει να εξεταστούν σε βάθος, και ως εκ τούτου η ανίχνευση των ακραίων τιμών πραγματοποιήθηκε στο σύνολο δεδομένων και οι επισημανθείσες ακραίες τιμές αφαιρέθηκαν από αυτό. Για την πρόβλεψη των τιμών, τα μοντέλα που εφαρμόστηκαν είναι η γραμμική και η λογιστική παλινδρόμηση και Random Forest. Μετά την εφαρμογή των προαναφερθέντων τριών προσεγγίσεων επιλέχθηκε το καλύτερο μοντέλο με βάση την τιμή RMSE. Στην παρούσα εργασία, η μικρότερη τιμή του RMSE ήταν για τη μέθοδο Random Forest.

Όλο αυτό έγινε, διότι με τη συνεχή αύξηση των καταχωρίσεων, η δημιουργία της σωστής καταχώρισης με σκοπό να παραμείνει στην αγορά δεν είναι μόνο επιτακτικό ζήτημα αλλά και ανταγωνιστικό. Αν και οι επισκέπτες αντιμετωπίζουν επίσης ένα δικό τους ζήτημα, το οποίο είναι η μη διαθεσιμότητα καταλύματος. Αυτό μπορεί να οφείλεται σε πολλούς λόγους, όπως εποχικότητα, απότομες ακυρώσεις κρατήσεων ή κάποια αλλαγή στις προτιμήσεις. Οι συγγραφείς της συγκεκριμένης έρευνας προτείνουν την ανάλυση των διαθέσιμων στο κοινό πληροφοριών καταχώρισης στην Airbnb τα τελευταία τρία χρόνια, για να προσπαθήσουν να επικυρώσουν τις υποθέσεις τους. Είναι πολύ σημαντικό για έναν ιδιοκτήτη ενός ή πολλών καταλυμάτων να βρει τη σωστή τιμή για την ιδιοκτησία του και η πλατφόρμα της Airbnb παρέχει γενικές οδηγίες ως προς τη σωστή τιμολόγηση. Οπότε, η τιμή που θα βάλει εξαρτάται από αυτόν.

Η συγκεκριμένη έρευνα περιλαμβάνει την ανάλυση του συνόλου των δεδομένων της Airbnb σε διάφορες πόλεις των Ηνωμένων Πολιτειών και την οπτικοποίηση των παρατηρήσεων που βρέθηκαν κατά την ανάλυση προκειμένου να βελτιωθεί το μάρκετινγκ και να κερδιθεί η ικανοποίηση των πελατών παρέχοντας τις απαιτούμενες διευκολύνσεις. Η εργασία επικεντρώνεται κυρίως στην πραγματοποίηση της κατάλληλης έρευνας σχετικά με τα δεδομένα της Airbnb που προέρχονται από διαφορετικές και πιο δημοφιλείς πόλεις (συνολικά 28) σε όλη

την επικράτεια των ΗΠΑ. Ο κύριος στόχος αυτής της έρευνας είναι να γίνουν προβλέψεις σχετικά με τις τιμές των κατοικιών Airbnb, συνδέοντας τις τιμές με τις τοποθεσίες και την ανάλυση των ελάχιστων και μέγιστων αριθμών των κρατήσεων ανά μήνα, έτσι, η έρευνα αυτή όχι μόνο βοηθάει τους πελάτες της Airbnb να ανταποκριθούν στα προτιμώμενα καταλύματά τους, αλλά επίσης βοηθάει και τους οικοδεσπότες της Airbnb αυξάνοντας τους πελάτες τους.

3.2.2. Χαρακτηριστικά και Μέθοδοι

Στη συγκεκριμένη έρευνα, οι ερευνητές χρησιμοποίησαν τις εξής μεθόδους:

1. Προ-επεξεργασία

Τα δεδομένα υπάρχουν, οπότε τώρα το επόμενο βήμα είναι να καθοριστεί εάν υπάρχει σημαντική σχέση μεταξύ των μεταβλητών στα δεδομένα. Μέχρι το συγκεκριμένο σημείο της έρευνας, οι συγγραφείς έχουν ανακαλύψει ότι η τιμή μιας καταχώρισης φαίνεται να επηρεάζεται από πολλούς παράγοντες. Αφού επιλέχθηκε το σύνολο των χαρακτηριστικών, οι συγγραφείς προσπάθησαν να αναπτύξουν ένα μοντέλο πρόβλεψης χρησιμοποιώντας την ανάλυση παλινδρόμησης. Αυτή η στατιστική τεχνική χρησιμοποιείται για τον προσδιορισμό της σχέσης μεταξύ μιας μεμονωμένης εξαρτημένης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Στη συγκεκριμένη έρευνα οι συγγραφείς αναφέρονται στη γραμμική παλινδρόμηση, τη λογιστική παλινδρόμηση και την ταξινόμηση Random Forest.

2. Λογιστική Παλινδρόμηση

Ο στόχος της γραμμικής παλινδρόμησης, είναι να εκτιμηθούν οι τιμές για τους συντελεστές των μοντέλων $m_1, m_2, m_3 \dots m_n$ και να γίνει προσπάθεια για την εκπαίδευση των μοντέλων με το μικρότερο μέσο τετραγωνικό σφάλμα με σκοπό να προβλεφθούν οι τιμές για τη στοχευμένη μεταβλητή. Παρόμοιες προβλέψεις γίνονται με τα μοντέλα λογιστικής παλινδρόμησης, αλλά με μία προσθήκη. Το μοντέλο λογιστικής παλινδρόμησης υπολογίζει ένα σταθμισμένο άθροισμα για τις ανεξάρτητες μεταβλητές όπως αυτές που πάρθηκαν στη γραμμική παλινδρόμηση αλλά αυτή τη φορά, η συνάρτηση που χρησιμοποιείται είναι μια μη γραμμική συνάρτηση γνωστή ως λογιστική ή σιγμοειδής και παράγει την εξαρτημένη μεταβλητή Y . Η σιγμοειδής / λογιστική συνάρτηση προσαρμόζεται στην ακόλουθη εξίσωση :

$$Y = 1 / (1 + e^{-1})$$

$$\text{όπου } Y = b_0 + (m_1 * x_1) + (m_2 * x_2) + (m_3 * x_3) \dots (m_n * x_n).$$

Η έξοδος ερμηνεύεται ως δυαδική τιμή, δηλαδή 0 ή 1. Ο σκοπός εδώ είναι να προβλεφθεί η τιμολόγηση μιας εισαγωγής, έτσι ώστε να δείχνει αν ενοικιαστεί ή όχι το κατάλυμα. Μερικές φορές, γίνεται προσπάθεια για να δούμε αν ένα δωμάτιο ενοικιάζεται ή όχι σε έναν επισκέπτη με βάση τον αριθμό των κριτικών που έχει η συγκεκριμένη καταχώριση, τη βαθμολογία των κριτικών, την ακρίβεια της βαθμολογίας των κριτικών, τη διαθεσιμότητα του δωματίου και

τους ελάχιστους και μέγιστους αριθμούς διανυκτερεύσεων που μπορεί να περάσει ένας επισκέπτης εκεί. Στο ουσία, μπορεί να χρησιμοποιηθεί ως στόχος, δηλαδή το μοντέλο προβλέπει 1, όταν η πιθανότητα της μεταβλητής είναι μεγαλύτερη από 0,5 και όταν η πιθανότητα είναι μικρότερη από 0,5 τότε προβλέπει 0.

3. Ανίχνευση ακραίων τιμών

Οι ακραίες τιμές σε ένα σύνολο δεδομένων είναι εκείνα τα σημεία δεδομένων που αποκλίνουν πολύ από τα κανονικά ή τα μέσα σημεία δεδομένων. Αυτά τα μη φυσιολογικά μοτίβα στο σύνολο των δεδομένων πρέπει να εξετάζονται προσεκτικά και να αφαιρούνται από το συγκεκριμένο σύνολο δεδομένων, προκειμένου να αυξηθεί η ακρίβεια του μοντέλου πρόβλεψης. Το ζήτημα που συνδέεται με αυτή την ιδέα είναι ότι δεν υπάρχουν διαθέσιμα επισημασμένα δεδομένα για να χρησιμοποιηθούν ως σύνολο δεδομένων εκπαίδευσης για την ανίχνευση ανωμαλιών, επομένως μια μη επιβλεπόμενη μηχανή αλγοριθμικής μάθησης απαιτείται για την προσέγγιση αυτή. Ως εκ τούτου, το "One Class Support Vector Machine" υλοποιείται. Ένα SVM είναι ένας αλγόριθμος μηχανικής μάθησης χωρίς επίβλεψη ο οποίος μπορεί να χρησιμοποιηθεί για την ταξινόμηση των ανώμαλων σημείων στα δεδομένα σε ένα σύνολο δεδομένων. Η μέθοδος One Class SVM δημιουργεί μια απόφαση με βάση τα δεδομένα και οτιδήποτε βρίσκεται έξω από το όριο ταξινομείται ως ακραίο σημείο. Η ανίχνευση των ακραίων σημείων είναι πολύ απαραίτητη για τη βελτίωση της πρόβλεψης. Στη συγκεκριμένη έρευνα, οι τιμές των καταχωρίσεων για την Airbnb μπορεί να περιέχουν αυτές τις ακραίες τιμές και επομένως αυτό μπορεί να επηρεάζει τη συνολική απόδοση του μοντέλου πρόβλεψης. Ως εκ τούτου, το SVM μίας κατηγορίας θα βοηθήσει στην ανίχνευση των ανωμαλιών ή των ακραίων τιμών στη συνεχή μεταβλητή τιμή και θα βελτιώσει την ακρίβεια του μοντέλου για την περαιτέρω ανάλυση πρόβλεψης για την τιμή.

3.2.3. Αποτελέσματα

Τα αποτελέσματα που έβγαλαν οι ερευνητές από το Πανεπιστήμιο της Καλιφόρνια, είχαν ως εξής:

1. Γραμμική Παλινδρόμηση

Το Σχήμα 3 δείχνει τη διαφορά των πραγματικών τιμών και των τιμών που προβλέπονται από το μοντέλο γραμμικής παλινδρόμησης. Μετά την εξέταση του γραφήματος, είμαστε σε θέση να αναλύσουμε τον τρόπο με τον οποίο η πραγματική τιμή μεταβάλλεται από τις προβλεπόμενες τιμές με τις οποίες μπορούμε να υπολογίσουμε το μέσο τετραγωνικό σφάλμα για να ελέγξουμε πόσο ακριβές είναι το μοντέλο για την πρόβλεψη των τιμών με βάση ορισμένα χαρακτηριστικά.



Σχήμα 3. Πραγματικές τιμές εναντίον των τιμών που προβλέπει το μοντέλο γραμμικής παλινδρόμησης.^[15]

2. Λογιστική Παλινδρόμηση

Για τη λογιστική παλινδρόμηση, αντί να σχεδιαστεί μια γραμμική σχέση μεταξύ των ανεξάρτητων και εξαρτημένων μεταβλητών, διαιρείται το σύνολο δεδομένων σε δεδομένα εκπαίδευσης και δεδομένα δοκιμής για να ληφθεί η λογιστική σχέση. Στη συγκεκριμένη έρευνα, χωρίζονται τα δεδομένα εκπαίδευσης και δοκιμής ως 80% και 20% αντίστοιχα. Σε αυτό το μοντέλο, πρώτα εκπαιδεύεται κάποιο μέρος των δεδομένων για το οποίο οι τιμές είναι ήδη δοσμένες. Στη συνέχεια, έχουμε μια λογιστική εξίσωση που αντιπροσωπεύει την πιθανότητα της τιμής. Αφού παρθεί η πιθανότητα των τιμών χρησιμοποιώντας τα εκπαιδευμένα δεδομένα, προβλέπονται οι τιμές για τα δεδομένα δοκιμής χρησιμοποιώντας τη λογιστική εξίσωση που βρέθηκε από τα δεδομένα εκπαίδευσης. Η ακρίβεια του ταξινομητή λογιστικής παλινδρόμησης στο σύνολο δοκιμής είναι 0,78.

3. Ταξινομητής Random Forest

Σε αυτό το μοντέλο, χρησιμοποιήθηκαν οι ελάχιστες διανυκτερεύσεις, οι μέγιστες διανυκτερεύσεις, η διαθεσιμότητα 365 ημερών, ο αριθμός των κριτικών, η βαθμολογία της κριτικής, η ακρίβεια της βαθμολογίας της κριτικής και τα καταλύματα για την πρόβλεψη της τιμής. Το συγκεκριμένο μοντέλο εκτίμησε τις τιμές μετά την προσαρμογή διαφόρων ταξινομητών δέντρων απόφασης σε διάφορες υποενοότητες του συνόλου δεδομένων. Η τελική απάντηση παρέχεται μετά τη μέση τιμή των προβλέψεων, προκειμένου να βελτιωθεί η ακρίβεια και η ισχύς κατά την προσαρμογή. Επιτεύχθηκε ακρίβεια 87% χρησιμοποιώντας το μοντέλο Random Forest.

3.2.4. Συμπεράσματα

Ο κύριος στόχος της συγκεκριμένης έρευνας ήταν να γίνουν προβλέψεις τιμών για την Airbnb χρησιμοποιώντας διαφορετικούς αλγόριθμους μηχανικής μάθησης. Το κίνητρό των ερευνητών ήταν επίσης να ανακαλύψουν ποιοι αλγόριθμοι μηχανικής μάθησης λειτουργούν καλύτερα για το σύνολο δεδομένων στην πρόβλεψη των τιμών με μέγιστη ακρίβεια. Οι ερευνητές διαπίστωσαν ότι το μοντέλο Random Forest λειτουργεί καλύτερα με αυτό το είδος δεδομένων. Η έρευνα θα παρέχει στον οικοδεσπότη τις καλύτερες πληροφορίες σύμφωνα με τις ιδιαίτερες ανάγκες του, όπως τοποθεσία, αξιολογήσεις και τιμές. Η ακρίβεια της πρόβλεψης μοντέλων μπορεί να βελτιωθεί περαιτέρω χρησιμοποιώντας τις κριτικές και τα χαρακτηριστικά των καταχωρίσεων. Έτσι, ένας οικοδεσπότης θα έχει τη δυνατότητα να τιμολογήσει καλύτερα το κατάλυμά του.

3.3. Προβλέψεις γειτονιάς και τιμών για τις καταχωρίσεις του Σαν Φρανσίσκο στην Airbnb.^[16]

3.3.1. Εισαγωγή

Σε μία επόμενη έρευνα που πραγματοποιήθηκε στο Stanford University και συγκεκριμένα από τους Emily Tang και Kunal Sangani, αναλύονται προβλέψεις γειτονιάς και τιμών για τις καταχωρίσεις που έγιναν στην πόλη του Σαν Φρανσίσκο στην Airbnb, όπου και ιδρύθηκε η εταιρεία, μέχρι τις 2 Νοεμβρίου του 2015.

Χρησιμοποιώντας το πλήρες σύνολο των καταχωρίσεων που είναι διαθέσιμες στο Σαν Φρανσίσκο από την ιστοσελίδα Inside Airbnb, οι συγγραφείς της συγκεκριμένης εργασίας αναπτύσσουν ταξινομητές που είναι σε θέση να προβλέψουν τη γειτονιά και την τιμή μιας καταχώρισης. Η πρόβλεψη της γειτονιάς μιας καταχώρισης δίνει πληροφορίες για τα πολιτιστικά στοιχεία που είναι ορατά μέσω κειμένου, εικόνas και ανέσεων που θα μπορούσαν να συνδέσουν γειτονιές μεταξύ τους και θα μπορούσαν ενδεχομένως να εφαρμοστούν για ένα σύστημα συστάσεων (π.χ. "Εάν απολαύσατε τη διαμονή σας στο Haight Ashbury, σας συνιστούμε να δοκιμάσετε την πλατεία Alamo!"), και προβλέποντας την κατάλληλη τιμολόγηση μιας καταχώρισης, ειδικά λαμβάνοντας υπόψη λειτουργίες μοναδικές για την οικονομία διαμοιρασμού του τρόπου με τον οποίο οι οικοδεσπότες περιγράφουν την προσφορά τους, μπορεί να είναι χρήσιμη στα εσωτερικά εργαλεία τιμολόγησης που προσφέρει η Airbnb στους οικοδεσπότες της.

Η ιστοσελίδα Inside Airbnb παρέχει δεδομένα για το σύνολο των διαθέσιμων καταχωρίσεων στο Σαν Φρανσίσκο. Εισάγονται χαρακτηριστικά κειμένου, χαρακτηριστικά εικόνas και άλλες πληροφορίες που σχετίζονται με κάθε καταχώριση σε έναν ταξινομητή SVM για να προβλεφθεί η γειτονιά και η τιμή της εκάστοτε καταχώρισης. Στις επόμενες ενότητες, εξηγείται το πλαίσιο της έρευνας, περιγράφεται το σύνολο δεδομένων και οι μέθοδοι για την εξαγωγή χαρακτηριστικών και την ταξινόμηση και παρουσιάζεται μια ανάλυση των αποτελεσμάτων.

3.3.2. Χαρακτηριστικά και Μέθοδοι

Οι ερευνητές Emily Tang και Kunal Sangani από το Stanford University, οι οποίοι ήταν υπεύθυνοι για την έρευνα στην πρόβλεψη γειτονιών και στην πρόβλεψη τιμών πάνω στην ιδρυτική πόλη της Airbnb, δηλαδή το Σαν Φρανσίσκο, χρησιμοποίησαν πέντε σύνολα χαρακτηριστικών, τα οποία είναι: (1) χαρακτηριστικά πληροφοριών καταχώρισης, (2) χαρακτηριστικά πολωνυμικών λέξεων, (3) πολωνυμικά χαρακτηριστικά κατηγορίας λέξεων, (4) χαρακτηριστικά συναισθήματος κειμένου και (5) οπτικά χαρακτηριστικά. Τα χαρακτηριστικά πληροφοριών καταχώρισης περιλαμβάνουν τον τύπο του ακινήτου (διαμέρισμα, κρεβάτι και πρωινό, κοιτώνας, σπίτι κ.λπ.), τον τύπο του κρεβατιού (πραγματικό

κρεβάτι, κρεβάτι με αέρα κ.λπ.), την πολιτική ακύρωσης του οικοδεσπότη και τον αριθμό των κρεβατιών, των υπνοδωματίων, των λουτρών και των επισκεπτών που φιλοξενούνται.

Με σκοπό να δημιουργηθούν τα χαρακτηριστικά των λέξεων, συγκεντρώνονται το όνομα, η περίληψη, ο χώρος, η περιγραφή, οι προσφερόμενες εμπειρίες κάθε καταχώρισης, σημειώσεις, και το βιογραφικό του οικοδεσπότη. Στη συνέχεια, χρησιμοποιείται το PorterStemmer του πακέτου NLTK για να προσδιοριστεί η ρίζα των λέξεων και επιλέγονται τα 1.000 στελέχη που εμφανίζονται στις περισσότερες καταχωρίσεις στο σύνολο δοκιμών.

Μια πιο προσεκτική ματιά στο σύνολο δεδομένων αποκαλύπτει ότι λέξεις όπως comfortable και cozy μπορεί να χρησιμοποιούνται εναλλακτικά μεταξύ των καταχωρίσεων. Με σκοπό να αντιμετωπιστεί αυτό, επιλέγονται 9 κατηγορίες λέξεων: people, nightlife, activities, style, accessibility, culture, nature, amenities, και comfort. Για παράδειγμα, η κλάση λέξεων "style" περιλαμβάνει τα στελέχη των λέξεων modern, brand, victorian, style, decor, gorgeous, marble, elegant, boutique και trendy. Τα πολωνυμικά χαρακτηριστικά της κλάσης λέξεων είναι μετρήσεις του αριθμού των λέξεων που ανήκουν σε κάθε μία από τις εννέα κατηγορίες λέξεων σε κάθε καταχώριση, με το άθροισμα των χαρακτηριστικών να είναι κανονικοποιημένο ώστε το άθροισμα να είναι ένα. Για τα χαρακτηριστικά συναισθήματος, χρησιμοποιείται το πακέτο TextBlob, το οποίο υπολογίζει την πολικότητα ενός τμήματος κειμένου υπολογίζοντας κατά μέσο όρο την πολικότητα κάθε λέξης στο κείμενο που περιλαμβάνεται στο λεξικό.

Τέλος, κατεβαίνουν όλες οι εικόνες που περιλαμβάνονται στην λίστα και εξάγονται οπτικά χαρακτηριστικά χρησιμοποιώντας το τυπικό μοντέλο bag of words στην όραση υπολογιστών. Επιλέγονται τυχαία δείγματα 100 εικόνων για να δημιουργηθεί ένα λεξικό 1.000 λέξεων. Για τη δημιουργία του λεξικού, εξάγονται Speeded Up Robust Features (SURF) περιγραφείς από τις 100 εικόνες χρησιμοποιώντας το OpenCV. Το SURF είναι ένας αποδοτικός και αμετάβλητος ανιχνευτής και περιγραφέας σημείων ενδιαφέροντος που χρησιμοποιείται για την εύρεση ομοιοτήτων μεταξύ εικόνων. Οι ερευνητές χρησιμοποιούν την ομαδοποίηση K-means για να σχηματίσουν 1.000 συστάδες ή οπτικές λέξεις για το λεξικό. Χρησιμοποιώντας αυτό το λεξικό οπτικών λέξεων, δημιουργούνται διανύσματα χαρακτηριστικών για κάθε καταχώριση εξάγοντας περιγραφικά στοιχεία SURF από την εικόνα της λίστας. Για κάθε περιγραφέα, επιλέγεται η πλησιέστερη ομάδα του λεξικού. Εν συντομία, το διάνυσμα οπτικών χαρακτηριστικών περιέχει τη συχνότητα κάθε οπτικής λέξης στην εικόνα της λίστας και είναι κανονικοποιημένο.

Τα παραπάνω χαρακτηριστικά εισάγονται σε μια Μηχανή Διανυσμάτων Υποστήριξης (SVM), που υλοποιείται με τη χρήση του πακέτου sklearn's SVM. Οι ερευνητές στη συνέχεια σχεδιάζουν καμπύλες μάθησης για κάθε ταξινομητή και παρατηρείται ότι το μοντέλο έχει υψηλή διακύμανση, γεγονός που υποδηλώνει πρόβλημα υπερπροσαρμογής. Για να αντιμετωπιστεί αυτό το πρόβλημα, χρησιμοποιείται Recursive Feature Elimination (RFE) του sklearn. Ο RFE είναι ένας αλγόριθμος που εξετάζει αναδρομικά μικρότερα υποσύνολα όλων

των χαρακτηριστικών, αφαιρώντας τα χαρακτηριστικά με τα μικρότερα βάρη μετά από εκπαίδευση σε ένα σύνολο χαρακτηριστικών σε κάθε επανάληψη.

3.3.3. Αποτελέσματα

Στην έρευνά τους οι Emily Tang και Kunal Sangani, όπου έγινε για τη γενέτειρα της Airbnb και συγκεκριμένα το Σαν Φρανσίσκο, διαπιστώνουν ότι το μοντέλο πρόβλεψης τιμών επιτυγχάνει ακρίβεια δοκιμής 81,2% χρησιμοποιώντας όλα τα χαρακτηριστικά, η οποία είναι σημαντικά υψηλότερη από την ακρίβεια πιθανότητας του 50%. Για το μοντέλο πρόβλεψης γειτονιάς, οι ερευνητές διαπιστώνουν ότι η υψηλότερη ακρίβεια που επιτυγχάνεται είναι 42,2% χρησιμοποιώντας μόνο τα χαρακτηριστικά κειμένου. Στην πρόβλεψη γειτονιάς, η βασική τιμή ακρίβειας της πιθανότητας είναι 8%.

Η επιλογή 800 χαρακτηριστικών για το RFE δίνει τις περισσότερες ακριβείς τιμές και η επιλογή 150 χαρακτηριστικών κάνει το ίδιο και για τις γειτονιές. Οι ερευνητές κατασκεύασαν τα μοντέλα, εκτέλεσαν το RFE χρησιμοποιώντας τον καλύτερο αριθμό χαρακτηριστικών και σχεδίασαν τη μάθηση στις καμπύλες μάθησης ξανά για να διαπιστώσουν αν τα προβλήματα υπερπροσαρμογής για τις γειτονιές και τις τιμές έχουν μειωθεί.

Συγκρίνοντας τις δύο καμπύλες μάθησης για το μοντέλο πρόβλεψης γειτονιάς, διαπιστώθηκαν παρόμοιες βελτιώσεις. Μετά την επιλογή 150 χαρακτηριστικών, η ακρίβεια εκπαίδευσης δεν ήταν πλέον μέγιστη στο 1 και μειώθηκε σε περίπου 0,8. Επιπλέον, διαπιστώθηκε ότι η επιλογή χαρακτηριστικών βελτιώνει την ακρίβεια ελαφρώς από 42,3% σε 43,1%. Τέλος, παρατηρήθηκε ότι το CVS αυξήθηκε καθώς αυξήθηκε και ο αριθμός των εκπαιδευτικών δειγμάτων. Αυτή η ανοδική τάση υποδηλώνει ότι, αν ήταν διαθέσιμα περισσότερα παραδείγματα εκπαίδευσης, μπορεί να είχαν βοηθήσει στη βελτίωση του ταξινομητή.

3.3.4. Συμπεράσματα

Η ανάλυση των δύο ερευνητών από το Stanford University για τις καταχωρίσεις στην πόλη του Σαν Φρανσίσκο, δείχνει ότι μπορούμε να προβλέψουμε με επιτυχία τις γειτονιές και το εύρος των τιμών χρησιμοποιώντας μια σειρά χαρακτηριστικών που εξάγονται από τις καταχωρίσεις. Αρχικά, οι ερευνητές ανακαλύπτουν ότι υπάρχει υψηλή διακύμανση στους ταξινομητές, αλλά μετά την επιλογή των κατάλληλων χαρακτηριστικών καταλήγουν στο συμπέρασμα ότι υπάρχει η δυνατότητα αυτά τα χαρακτηριστικά να αξιοποιηθούν με σωστό τρόπο για την πρόβλεψη της γειτονιάς και της τιμής.

Τέλος, παρατηρείται ότι υπάρχει σημαντική επικάλυψη μεταξύ των χαρακτηριστικών που προβλέπουν τη γειτονιά και την τιμή. Σε μελλοντικές εργασίες, θα ήταν ενδιαφέρον να κατανοηθούν βαθύτερα οι συσχετίσεις μεταξύ γειτονιών και τιμών. Οι αριστοκρατικές

γειτονίες έχουν φυσικά υψηλότερες τιμές στις καταχωρίσεις και η ανίχνευση των χαρακτηριστικών των γειτονιών που ενθαρρύνουν τους καταναλωτές να πληρώνουν υψηλότερες τιμές μπορεί να αποτελέσει σημαντική εφαρμογή για την κτηματομεσιτική και την πολεοδομική νομοθεσία.

3.4. Πρόβλεψη των τιμών των καταχωρίσεων της Airbnb σε διαφορετικές πόλεις.

[17]

3.4.1. Εισαγωγή

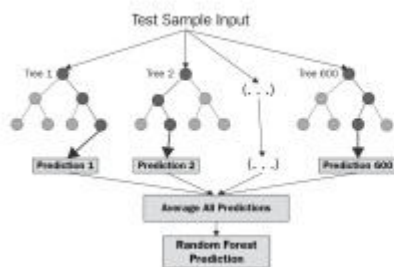
Στο κοντινό 2019, οι Yuanhang Luo, Xuanyu Zhou και Yulian Zhou, έκαναν μία έρευνα σχετικά με την πρόβλεψη τιμών στις καταχωρίσεις της Airbnb για διάφορες πόλεις σε όλο τον κόσμο και πιο συγκεκριμένα για τη Νέα Υόρκη, το Παρίσι και το Βερολίνο με διάφορες τεχνικές μηχανικής μάθησης.

Σε αντίθεση με τα ξενοδοχεία, τα οποία έχουν το δικό τους σύστημα τιμολόγησης, οι τιμές του Airbnb καθορίζονται συνήθως από τους οικοδεσπότες εμπειρικά. Έτσι, υπάρχουν προκλήσεις για τους νέους οικοδεσπότες, καθώς και για τους υπάρχοντες οικοδεσπότες με νέες καταχωρίσεις, ώστε να καθορίσουν τις τιμές σε λογικά πλαίσια χωρίς όμως να χάσουν τη δημοτικότητά τους. Από την πλευρά των καταναλωτών, όμως, οι οποίοι μπορούν να συγκρίνουν την τιμή σε άλλες παρόμοιες καταχωρίσεις, εξακολουθεί να είναι πολύτιμο γι' αυτούς να γνωρίζουν αν η τρέχουσα τιμή είναι αξιόλογη και αν είναι καλή στιγμή για να κάνουν κράτηση των δωματίων.

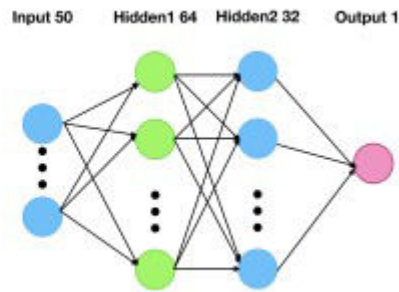
Οι νυχτερινές τιμές για ενοικίαση στην Airbnb εξαρτάται από πολλούς παράγοντες και οι ερευνητές χωρίζουν τον τύπο εισόδου σε 4 κατηγορίες, συμπεριλαμβανομένων των συνεχών χαρακτηριστικών, των κατηγορικών χαρακτηριστικών, των χαρακτηριστικών κειμένου και της ημερομηνίας. Εξήχθησαν περισσότερα από 60 χαρακτηριστικά από το σύνολο των δεδομένων. Στη συνέχεια παρατίθενται μόνο μερικά από αυτά που είναι σημαντικά για την εργασία, όπως το μέγεθος του δωματίου, επιπλέον χρεώσεις, βαθμολογία κριτικών, τοποθεσία, εγκαταστάσεις, και σχετικά χαρακτηριστικά με την κράτηση. Χρησιμοποιείται γραμμική παλινδρόμηση, παλινδρόμηση K κοντινότερων γειτόνων, παλινδρόμηση Random Forest, XGBoost και νευρωνικά δίκτυα, για την πρόβλεψη των τιμών.

3.4.2. Χαρακτηριστικά και Μέθοδοι

Οι ερευνητές Yuanhang Luo, Xuanyu Zhou και Yulian Zhou, που ασχολήθηκαν με μία έρευνα σχετικά με την πρόβλεψη τιμών στις καταχωρίσεις της Airbnb για διάφορες πόλεις σε όλο τον κόσμο και πιο συγκεκριμένα για τη Νέα Υόρκη, το Παρίσι και το Βερολίνο, κινήθηκαν με μεθόδους όπως γραμμική παλινδρόμηση με κανονικοποίηση L2, KNN, Random Forest, XGBoost και Νευρωνικά Δίκτυα. Η γραμμική παλινδρόμηση χρησιμοποιήθηκε με σκοπό να μειωθεί η συνάρτηση κόστους. Η KNN χρησιμοποιήθηκε με σκοπό να βρεθεί ο μέσος όρος των τιμών της πρόβλεψης. Το Random Forest, βοήθησε ώστε να προβλεφθεί η τελική έξοδος από συγκεκριμένα χαρακτηριστικά και το XGBoost βοήθησε ώστε να φαίνεται το αποτέλεσμα καθ' όλη τη διάρκεια του ελέγχου.



Σχήμα 4. Απεικόνιση της παλινδρόμησης Random Forest.^[27]



Σχήμα 5. Απεικόνιση των Νευρωνικών Δικτύων.^[17]

3.4.3. Αποτελέσματα

Οι τρεις ερευνητές χρησιμοποίησαν τις προεπιλεγμένες ρυθμίσεις για όλες τις προσεγγίσεις μηχανικής μάθησης που εισήχθησαν από το sklearn. Πειραματίστηκαν με διαφορετικές παραμέτρους, όπως η ισχύς της κανονικοποίησης, ο ρυθμός μάθησης, το k στο KNN κ.λπ. και οι μετρικές εξόδου παρέμειναν σε μεγάλο βαθμό αμετάβλητες.

Λόγω της ύπαρξης ασυνήθιστα υψηλών τιμών στις καταχωρίσεις, η άμεση πρόβλεψη στις τιμές παρουσίασε συμβιβασμένη απόδοση. Έτσι, οι ερευνητές έβαλαν όριο τιμής τα 500 δολάρια δοκίμασαν διαφορετικά μοντέλα στο dataset της Νέας Υόρκης μόνο με τα συνεχή χαρακτηριστικά. Το XGBoost και τα νευρωνικά δίκτυα πέτυχαν την καλύτερη απόδοση.

Στη συνέχεια, οι ερευνητές εξέτασαν περαιτέρω τις επιδόσεις του μοντέλου XGBoost χωρίς αλλά και με διαφορετικές ετικέτες και διαπίστωσαν ότι ο λογαριθμικός μετασχηματισμός βελτιώνει σε μεγάλο βαθμό την απόδοση του μοντέλου ακόμη και χωρίς όριο τιμών, όπως φαίνεται στον Πίνακα 3. Μετά την προσθήκη των όρων αλληλεπίδρασης στα συνεχή χαρακτηριστικά, η τιμή του r^2 αυξήθηκε από 0,669 σε 0,676.

Model	Feature	Label	Train		Test	
			MSE	r-squared	MSE	r-squared
Linear regression	C ^a	TH ^b	3514.64	0.489	3541.99	0.497
K-nearest neighbor	C	TH	4466.30	0.351	4760.61	0.323
Random forest	C	TH	1706.86	0.752	2427.05	0.655
Neural network	C	TH	2059.43	0.702	2367.44	0.665
XGBoost	C	TH	2273.98	0.669	2357.83	0.665

^aContinuous feature with normalization

^bThreshold $y \leq 500$

Πίνακας 2. Συνεχή χαρακτηριστικά με διαφορετικά μοντέλα.^[17]

			Train		Test	
Model	Feature	Label	MSE	r-squared	MSE	r-squared
XGBoost	C	N/A	16748.6	0.558	58003.1	0.195
XGBoost	C	Sqrt ^a	7.17806	0.652	10.4435	0.537
XGBoost	C	Log ^b	0.13905	0.695	0.14765	0.669
XGBoost	C+I ^c	Log	0.13061	0.708	0.14704	0.676

^a $y' = \sqrt{y}$

^b $y' = \log y$

^cAdd 2-degree interaction

Πίνακας 3. Το μοντέλο XGBoost με διαφορετικούς μετασχηματισμούς ετικετών.^[17]

3.4.4. Συμπεράσματα

Τέλος, οι τρεις ερευνητές πραγματοποίησαν αρκετούς πειραματισμούς στα χαρακτηριστικά και πειραματίστηκαν με διάφορες προσεγγίσεις μηχανικής μάθησης για την πρόβλεψη των τιμών καταχώρισης στην Airbnb. Κατέληξαν στο συμπέρασμα ότι το XGBoost και το νευρωνικό δίκτυο υπερτερούν έναντι άλλων διαφορετικών προσεγγίσεων και πετυχαίνουν τιμές μεγαλύτερες του 0,7 στο r^2 . Συμπεράναν επίσης, ότι εκπαιδεύοντας το νευρωνικό δίκτυο με διάφορα datasets από την πόλη της Νέας Υόρκης, του Παρισιού και του Βερολίνου, αυξάνονται τα γενικά χαρακτηριστικά και καταστέλλονται τα ειδικά χαρακτηριστικά των πόλεων. Αυτό το αποτέλεσμα παρουσιάζει ιδιαίτερο ενδιαφέρον, καθώς υποδεικνύει τη δυνατότητα εκπαίδευσης ενός γενικευμένου μοντέλου σε σύνολα δεδομένων από πολλές πόλεις.

3.5. Learning-based μοντέλο πρόβλεψης τιμών για την Airbnb.^[18]

3.5.1. Εισαγωγή

Φέτος, το 2021, ο ερευνητής Siqi Yang από την Κίνα, δημιούργησε ένα learning-based μοντέλο πρόβλεψης με βάση την αγορά στην Κίνα. Το τρέχον μερίδιο αγοράς της Airbnb είναι μόλις 8,7% στην Κίνα, δηλαδή πολύ χαμηλότερο από άλλες διαδικτυακές πλατφόρμες. Για να επεκτείνει το μερίδιο αγοράς της, η πλατφόρμα της Airbnb θα πρέπει να βοηθήσει τους οικοδεσπότες να καθορίζουν αποτελεσματικά προσαρμοσμένες τιμές, δηλαδή η τιμή να μπορεί να γίνει αποδεκτή και από τους οικοδεσπότες αλλά ταυτόχρονα και από τους επισκέπτες. Η σωστή τιμολόγηση, σύμφωνα με τον Siqi Yang, μπορούν να βοηθήσουν τους οικοδεσπότες να αποκτήσουν περισσότερα κέρδη και να προσελκύσουν περισσότερους επισκέπτες με σκοπό να κάνουν κράτηση μέσω της Airbnb αντί για ένα ξενοδοχείο ή άλλες πλατφόρμες, το οποίο είναι ένα κρίσιμο και δύσκολο έργο της πλατφόρμας της Airbnb.

Παρά το γεγονός ότι η Airbnb προσφέρει εργαλεία σύστασης τιμών με βάση τις ιδιότητες των κατοικιών της από το 2012 και αύξησε την ακρίβεια και την ευελιξία του μοντέλου το 2015, ο αλγόριθμος πρόβλεψης τιμών εξακολουθεί να χρειάζεται βελτίωση. Εν τω μεταξύ, οι περισσότερες μελέτες επικεντρώνονται σε πόλεις των Ηνωμένων Πολιτειών ή της Ευρώπης και όχι στην Κίνα, αγνοώντας τη μεγαλύτερη αγορά της Airbnb. Για να υπάρξει μια βαθύτερη κατανόηση του εξατομικευμένου μοντέλου τιμολόγησης της Airbnb, η συγκεκριμένη έρευνα επικεντρώνεται στην πόλη του Πεκίνου.

Τα χαρακτηριστικά τιμών της πλατφόρμας Airbnb έχουν δώσει ιδιαίτερη προσοχή στην "Εξυπνη Τιμολόγηση", η οποία επιτρέπει στους οικοδεσπότες να ορίζουν μία ελάχιστη και μία μέγιστη τιμή για τη δυναμική προσαρμογή των τιμών με βάση τις καμπύλες ζήτησης. Ο μηχανισμός "Εξυπνης Τιμολόγησης" λαμβάνει υπόψη χρονικές πληροφορίες, όπως τα Σαββατοκύριακα και οι αργίες. Επίσης, για μακροχρόνιες κρατήσεις, όπως εβδομάδες και μήνες, παρέχει "Συμβουλές Τιμών". Αυτές οι προτάσεις απευθύνονται σε οικοδεσπότες που σκοπεύουν να δημοσιοποιήσουν το σπίτι τους στο διαδίκτυο. Οι περισσότεροι οικοδεσπότες, από την άλλη πλευρά, δεν μοιράζονται πλέον απλώς τα σπίτια τους. Αγοράζουν ή νοικιάζουν ακίνητα και τα διακοσμούν προκειμένου να επωφεληθούν από το Airbnb μακροπρόθεσμα. Οι οικοδεσπότες χρειάζονται περισσότερες προτάσεις για την τοποθεσία, τη διακόσμηση, τον εξοπλισμό και άλλους παράγοντες.

Η παρούσα έρευνα επιδιώκει να κατασκευάσει ένα αξιόπιστο μοντέλο πρόβλεψης τιμών με τη χρήση μηχανικής μάθησης για να βοηθήσει τους οικοδεσπότες με την εκτίμηση των τιμών και τις προτάσεις σπιτιών, προκειμένου να αντιμετωπιστούν οι προαναφερθείσες δυσκολίες.

3.5.2. Χαρακτηριστικά και Μέθοδοι

Στις ίδιες κατευθύνσεις κινήθηκε κι ο Siqi Yang, ο οποίος δημιούργησε ένα learning-based μοντέλο πρόβλεψης με βάση την αγορά στην Κίνα και πιο συγκεκριμένα στο Πεκίνο. Ο Yang χρησιμοποίησε dataset από την ιστοσελίδα Inside Airbnb. Το σύνολο δεδομένων αυτό όμως έχει κάποιους περιορισμούς. Το πρώτο και κυριότερο πρόβλημα είναι οι ακραίες τιμές. Ορισμένες τιμές είναι παράλογες και πολύ υψηλές (π.χ. η υψηλότερη τιμή είναι πάνω από 70.000 yuan και ορισμένα κοινόχρηστα δωμάτια έχουν υψηλότερες τιμές από 5.000 yuan, ενώ η μέση τιμή είναι περίπου 300 yuan). Για να διαπιστωθούν οι λόγοι, είναι απαραίτητη μια έρευνα φτιαγμένη για τους οικοδεσπότες. Σύμφωνα με τις απαντήσεις τους, ορισμένες ασυνήθιστες τιμές παρουσιάζουν "δοκιμαστικές καταχωρίσεις", πράγμα που σημαίνει ότι η καταχώριση γίνεται μόνο για να δοκιμαστεί αν η λειτουργία του Airbnb λειτουργεί και ο οικοδεσπότης δεν έχει σκοπό να το νοικιάσει καθόλου βραχυπρόθεσμα. Επίσης, οι περισσότερες από τις καταχωρίσεις ασυνήθιστων τιμών έχουν υψηλές "ελάχιστες διανυκτερεύσεις", συνήθως πάνω από 30 ημέρες. Επειδή αυτοί οι οικοδεσπότες θα νοικιάσουν το ακίνητό τους μόνο μηνιαία.

Μετά την προεπεξεργασία των χαρακτηριστικών, από τα συνολικά 106 χαρακτηριστικά, διαγράφηκαν τα 62 και έμειναν 44. Ο Yang χρησιμοποίησε τις μεθόδους XGBoost και Νευρωνικά Δίκτυα.

3.5.3. Αποτελέσματα

Στην έρευνα του Siqi Yang, ο οποίος δημιούργησε ένα learning-based μοντέλο για το Πεκίνο της Κίνας, χρησιμοποιούνται το Mean Squared Error (MSE) και το R2 για να αξιολογηθούν τα εκπαιδευμένα μοντέλα. Επίσης, τα αποτελέσματα και τα χαρακτηριστικά του μοντέλου XGBoost μετριοούνται και παρουσιάζονται σε γραφήματα.

Πάντοτε, πριν από τη δημιουργία μοντέλων νευρωνικών δικτύων, χρησιμοποιείται πρώτα ένα μοντέλο XGBoost για την πρόβλεψη της τιμής, προκειμένου να δημιουργηθεί μια γραμμή βάσης για το επίπεδο ακρίβειας και να μετρηθεί η σημασία των χαρακτηριστικών.

Περίπου το ένα τρίτο των χαρακτηριστικών έχει τιμή σημαντικότητας κοντά στο 0 στο μοντέλο παλινδρόμησης XGBoost, που σημαίνει ότι θα μπορούσε ενδεχομένως να αφαιρεθεί. Ο Πίνακας 4. και ο Πίνακας 5. δείχνουν τα 5 σημαντικότερα και τα 5 λιγότερο σημαντικά χαρακτηριστικά.

Δεν αποτελεί έκπληξη το γεγονός ότι το πόσους ανθρώπους φιλοξενεί το ακίνητο είναι το πιο σημαντικό, ενώ είναι ένα από τα κύρια πράγματα που θα εξετάσουν οι άνθρωποι σε πρώτη φάση. Επίσης, οι τύποι δωματίων του ακινήτου και ορισμένες ανέσεις είναι εξίσου σημαντικοί.

Model Name: XGBoost		
	MSE	R^2
Train split	0.0882	0.8112
Test split	0.1648	0.6549

Πίνακας 4. Αποτελέσματα του μοντέλου XGBoost.^[18]

Feature Importance Top 5		
	Feature	weight
1	accommodates	0.156459
2	room type: Entire home/apt	0.065606
3	bathrooms	0.043361
4	tv	0.036081
5	breakfast	0.031836

Πίνακας 5. Τα 5 πιο σημαντικά χαρακτηριστικά.^[18]

Feature Importance Bottom 5		
	Feature	weight
1	neighborhood cleansed	0.000000
2	review scores communication 0.8/10	0.000000
3	review scores checking 10/10	0.001563
4	time science last review 2-6 months	0.001717
5	review scores accuracy 9/10	0.001747

Πίνακας 6. Τα 5 λιγότερο σημαντικά χαρακτηριστικά.^[18]

Η ελάχιστη σημασία των αξιολογήσεων από την αγορά του Πεκίνου είναι πρωτοφανής. Σύμφωνα με τα στοιχεία της Airbnb για το Πεκίνο, οι επισκέπτες στο Πεκίνο δίνουν μεγαλύτερη αξία στους τύπους ακινήτων, τις ανέσεις και την τοποθεσία από ό,τι στις αξιολογήσεις.

Ο Siqi Yang, στην έρευνά του κατέγραψε και ένα νευρωνικό δίκτυο τριών επιπέδων, το οποίο χρησιμοποιεί μια συνάρτηση ενεργοποίησης κανόνα για το κρυφό στρώμα και μια γραμμική συνάρτηση ενεργοποίησης για το επίπεδο εξόδου και κατέληξε στο συμπέρασμα ότι έχει χειρότερες επιδόσεις από το μοντέλο XGBoost.

Model Name: Neural Network		
	MSE	R^2 Score
Train_split	0.2296	0.5087
Test_split	0.2379	0.5019

Πίνακας 7. Αποτελέσματα του Νευρωνικού Δικτύου.^[18]

3.5.4. Συμπεράσματα

Ο Siqu Yang, στη συγκεκριμένη έρευνα για την πόλη του Πεκίνου της Κίνας, προσπάθησε να καταλήξει στην καλύτερη απόδοση του μοντέλου πρόβλεψης τιμών για την Airbnb μέσω μηχανικής μάθησης, με σκοπό να βοηθήσει τους οικοδεσπότες της Airbnb να διαμορφώσουν τις κατάλληλες τιμές για τα καταλύματά τους. Χρησιμοποίησε τεχνικές μηχανικής μάθησης όπως tree-based XGBoost και νευρωνικά δίκτυα. Ο ερευνητής καταλήγει στο συμπέρασμα πως το μοντέλο XGBoost αποδίδει καλύτερα και έτσι επιλέγεται ως το τελικό μοντέλο πρόβλεψης τιμών. Επίσης, προτείνονται κάποιες μέθοδοι ώστε οι οικοδεσπότες να αυξήσουν τις προτεινόμενες τιμές, όπως η έγκαιρη προετοιμασία των καταχωρίσεων και η διακόσμησή τους. Επιπλέον, εξετάζονται οι περίεργες τιμές καταχωρίσεων στη συγκεκριμένη πόλη λόγω του ότι υπάρχουν καταχωρίσεις με μηνιαίο ενοίκιο. Αποδεικνύεται ότι οι ακραίες τιμές έχουν σημαντική επίδραση στην ακρίβεια του μοντέλου πρόβλεψης τιμών.

3.6. Γενικά συμπεράσματα και μελλοντικές κατευθύνσεις

Στην πρώτη έρευνα που έγινε, οι επιστήμονες προσπάθησαν να βρουν το μοντέλο με τις καλύτερες επιδόσεις για την πρόβλεψη τιμών στην Airbnb, με διάφορες τεχνικές μηχανικής μάθησης, όπως γραμμική παλινδρόμηση, tree-based μοντέλα, SVR και νευρωνικά δίκτυα, μαζί με αναλύσεις των χαρακτηριστικών που αναλύθηκαν με σκοπό την επίτευξη των καλύτερων αποτελεσμάτων όσον αφορά το Mean Squared Error (MSE), το Mean Absolute Error (MAE) και το R^2 . Οι ερευνητές ανακάλυψαν ότι μεταξύ των μοντέλων που δοκιμάστηκαν, το Support Vector Regression (SVR) απέδωσε καλύτερα από όλα.

Στις ίδιες κατευθύνσεις κινήθηκαν και οι επιστήμονες της δεύτερης έρευνας, αφού το κίνητρό τους ήταν να ανακαλύψουν ποιοι αλγόριθμοι μηχανικής μάθησης λειτουργούν καλύτερα για το σύνολο δεδομένων στην πρόβλεψη των τιμών με μέγιστη ακρίβεια. Διαπίστωσαν όμως ότι το μοντέλο Random Forest λειτουργεί περισσότερο αποδοτικά από τα υπόλοιπα με το συγκεκριμένο dataset.

Στην επόμενη έρευνα που αναλύθηκε, οι δύο ερευνητές από το Stanford University ασχολήθηκαν με την επιτυχή πρόβλεψη των γειτονιών και το εύρος των τιμών χρησιμοποιώντας μια σειρά χαρακτηριστικών τα οποία εξήχθησαν από τις καταχωρίσεις.

Παρατηρούν επίσης ότι υπάρχει σημαντική επικάλυψη μεταξύ των χαρακτηριστικών που προβλέπουν τη γειτονιά και την τιμή. Σκοπός των δύο ερευνητών είναι να δημιουργήσουν στο μέλλον το δικό τους μοντέλο πρόβλεψης τιμών.

Σε διαφορετική κατεύθυνση από την άλλη, κινήθηκαν οι τρεις ερευνητές της έρευνας που είχε να κάνει με την πρόβλεψη των τιμών των καταχωρίσεων της Airbnb σε διαφορετικές πόλεις και συγκεκριμένα στη Νέα Υόρκη, στο Παρίσι και στο Βερολίνο. Πραγματοποίησαν αρκετά πειράματα με διάφορα χαρακτηριστικά και αλγορίθμους και κατέληξαν στο συμπέρασμα ότι το μοντέλο XGBoost και το νευρωνικό δίκτυο υπερτερούν έναντι άλλων προσεγγίσεων.

Ο ερευνητής Siqi Yang στην έρευνά του για την πόλη του Πεκίνου της Κίνας, προσπάθησε επίσης να καταλήξει στην καλύτερη απόδοση του μοντέλου πρόβλεψης τιμών για την Airbnb μέσω μηχανικής μάθησης. Χρησιμοποίησε τεχνικές όπως tree-based XGBoost και νευρωνικά δίκτυα και κατέληξε στο συμπέρασμα πως το μοντέλο XGBoost αποδίδει καλύτερα.

Τέλος, φαίνεται πως υπάρχει πληθώρα ερευνών και υπάρχουν διάφορες μέθοδοι που μπορούν να χρησιμοποιηθούν για τη βελτίωση της πρόβλεψης τιμών στην κοινότητα της Airbnb, με σκοπό να επωφεληθούν και οι οικοδεσπότες αλλά και οι φιλοξενούμενοι. Οι μελλοντικές έρευνες θα μπορούσαν να μελετάνε άλλα σχήματα επιλογής χαρακτηριστικών, όπως η σημασία χαρακτηριστικών με Random Forest και η επιλογή τους με βάση τη συσχέτιση. Θα μπορούσαν επίσης να γίνουν περισσότερα πειράματα με αρχιτεκτονικές νευρωνικών δικτύων και να χρησιμοποιηθούν διαφορετικά μοντέλα μηχανικής μάθησης με εξειδικευμένο υλικό ώστε να είναι ενισχυμένα. Επιπλέον, στις μελλοντικές τους έρευνες πάνω στην πρόβλεψη τιμών, οι ερευνητές θα μπορούσαν να λάβουν περισσότερα παραδείγματα εκπαίδευσης από άλλες υπηρεσίες, όπως η Trivago, η VRBO κ.λπ.

4. ΑΝΑΖΗΤΗΣΗ

4.1. Εφαρμόζοντας βαθιά μάθηση στην αναζήτηση της Airbnb.^[19]

4.1.1. Εισαγωγή

Η συγκεκριμένη έρευνα, δημοσιεύθηκε για πρώτη φορά τον Αύγουστο του 2019 στην Αμερική από 11 επιστήμονες οι οποίοι ασχολήθηκαν με την εφαρμογή του deep learning στην αναζήτηση της Airbnb. Η εφαρμογή για την κατάταξη αναζήτησης είναι μία από τις μεγαλύτερες ιστορίες επιτυχίας της μηχανικής μάθησης για την Airbnb. Ένα μεγάλο μέρος των αρχικών κερδών υπήρξε εξαιτίας ενός ενισχυμένου μοντέλου δέντρου αποφάσεων. Τα κέρδη, ωστόσο, έμειναν στάσιμα με την πάροδο του χρόνου. Η συγκεκριμένη έρευνα εξετάζει το έργο που γίνεται στην εφαρμογή νευρωνικών δικτύων, σε μια προσπάθεια να αποφευχθεί αυτή η στασιμότητα.

Η πρώτη εφαρμογή της αναζήτησης με κατάταξη ήταν μία αυτόματα κατασκευασμένη συνάρτηση βαθμολόγησης. Η αντικατάσταση αυτής της συνάρτησης βαθμολόγησης με ένα μοντέλο δέντρου αποφάσεων με ενίσχυση διαβάθμισης (GBDT), προσέφερε στην Airbnb μία από τις μεγαλύτερες βελτιώσεις στις κρατήσεις κατοικιών στην ιστορία της.

Στη συγκεκριμένη έρευνα χρησιμοποιείται το μοντέλο των νευρωνικών δικτύων (NNs). Το υπό συζήτηση μοντέλο κατάταξης αναζήτησης αποτελεί μέρος ενός οικοσυστήματος μοντέλων, τα οποία συμβάλλουν στην απόφαση για το ποιες καταχωρίσεις θα παρουσιάζονται στον επισκέπτη. Αυτά περιλαμβάνουν μοντέλα που προβλέπουν την πιθανότητα ένας οικοδεσπότης να αποδεχθεί το αίτημα του επισκέπτη για κράτηση, μοντέλα που προβλέπουν την πιθανότητα ο επισκέπτης να βαθμολογήσει την ταξιδιωτική του εμπειρία με 5 αστέρια κλπ. Η έρευνα επικεντρώνεται σε ένα συγκεκριμένο μοντέλο σε αυτό το οικοσύστημα, το οποίο θεωρείται το πιο περίπλοκο κομμάτι του παζλ κατάταξης. Είναι υπεύθυνο για τις καταχωρίσεις που διατίθενται ανάλογα με την πιθανότητα κράτησης του επισκέπτη.

Είναι σύνηθες για τους επισκέπτες να κάνουν πολλές αναζητήσεις, κάνοντας κλικ σε ορισμένες καταχωρήσεις για να δουν τις λεπτομέρειες. Οι επιτυχημένες συνεδρίες ολοκληρώνονται με την κράτηση μιας καταχώρησης από τον επισκέπτη. Οι αναζητήσεις που πραγματοποιούνται από τον επισκέπτη και οι αλληλεπιδράσεις τους καταγράφονται. Κατά την εκπαίδευση του μοντέλου, ένα νέο μοντέλο έχει πρόσβαση στα αρχεία καταγραφής του τρέχοντος και των προηγούμενων μοντέλων που χρησιμοποιούνται στην παραγωγή. Το καινούριο μοντέλο εκπαιδεύεται ώστε να μάθει μια συνάρτηση βαθμολόγησης η οποία αποδίδει τις εντυπώσεις των καταχωρήσεων με κράτηση στα αρχεία καταγραφής σε όσο το δυνατόν υψηλότερη θέση. Στη συνέχεια, το καινούριο μοντέλο δοκιμάζεται online σε ένα πλαίσιο δοκιμών A/B, ώστε να διαπιστωθεί αν μπορεί να επιτευχθεί στατιστικά σημαντική αύξηση των μετατροπών σε σύγκριση με το τρέχον μοντέλο.

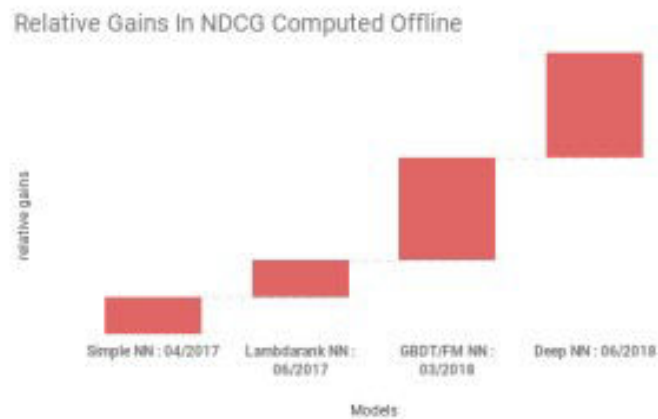
4.1.2. Χαρακτηριστικά και Μέθοδοι

Οι ερευνητές κινήθηκαν κυρίως γύρω από τα νευρωνικά δίκτυα και συγκεκριμένα χρησιμοποίησαν Simple NN, Lambdarank NN, Decision Tree / Factorization Machine NN και Deep NN. Η πραγματικότητα όμως είναι γεμάτη από αποτυχημένες προσπάθειες οι οποίες ξεπερνούν τις επιτυχημένες. Η ανάλυση κάθε αποτυχημένης προσπάθειας θα ήταν αρκετά χρονοβόρα για τη συγγραφή της έρευνας, για αυτό οι ερευνητές αποφάσισαν να αναφερθούν σε δύο αποτυχημένες προσπάθειες που αξίζουν αναφορά. Τα συγκεκριμένα μοντέλα είναι ενδιαφέροντα επειδή δείχνουν πώς κάποια τεχνική που είναι αρκετά αποτελεσματική και δημοφιλής, μπορεί να μη λειτουργήσει καλά σε συγκεκριμένους τομείς και να υπάρξουν προβλήματα. Αυτά τα μοντέλα είναι το Listing ID και το Multi-task learning, τα οποία θα αναλυθούν στην ενότητα των αποτελεσμάτων.

Στο πιο πρακτικό κομμάτι, οι αναζητήσεις ενός επισκέπτη εισχωρούν σε έναν διακομιστή Java ο οποίος κάνει ανάκτηση και βαθμολόγηση. Ο διακομιστής επίσης, παράγει αρχεία καταγραφής τα οποία αποθηκεύονται ως σειριοποιημένες περιπτώσεις Thrift. Τα αρχεία καταγραφής επεξεργάζονται με τη χρήση ενός Spark Pipeline με σκοπό τη δημιουργία δεδομένων εκπαίδευσης. Στη συνέχεια, η εκπαίδευση του μοντέλου γίνεται με τη χρήση του TensorFlow. Διάφορα εργαλεία γραμμένα σε Scala και Java χρησιμοποιούνται για την αξιολόγηση των μοντέλων και τον υπολογισμό μετρήσεων εκτός σύνδεσης. Στη συνέχεια, τα μοντέλα μεταφορτώνονται στον διακομιστή Java ο οποίος κάνει ανάκτηση και βαθμολόγηση. Όλα αυτά τα στοιχεία εκτελούνται σε AWS.

4.1.3. Αποτελέσματα

Η μετάβαση στη βαθιά μάθηση δεν ήταν το αποτέλεσμα μιας ατομικής κίνησης, αλλά το αποκορύφωμα μιας σειράς επαναληπτικών βελτιώσεων. Το Σχήμα 6α μας δείχνει τις συγκριτικές βελτιώσεις στο κύριο μετρικό NDCG εκτός σύνδεσης, όπου η καταχώριση της κλειστής λίστας έχει συνάφεια 1 και όλες οι άλλες καταχωρίσεις έχουν συνάφεια 0. Το NDCG (μειωμένο σωρευτικό κέρδος) είναι ένα μέτρο της ποιότητας κατάταξης. Στην ανάκτηση πληροφοριών, χρησιμοποιείται συχνά για τη μέτρηση της αποτελεσματικότητας των αλγορίθμων μηχανών αναζήτησης ή σχετικών εφαρμογών. Ο άξονας x απεικονίζει τα μοντέλα μαζί με το χρόνο κυκλοφορίας τους στην παραγωγή. Το Σχήμα 6β μας δείχνει τη συγκριτική αύξηση των μετατροπών από τα μοντέλα. Στο σύνολο, αυτό αντιπροσωπεύει μία από τις πιο επιδραστικές εφαρμογές της μηχανικής μάθησης στην Airbnb.



Σχήμα 6α. NDCG κέρδη εκτός σύνδεσης.^[19]



Σχήμα 6β. Κέρδη στις online κρατήσεις.^[19]

Από την άλλη, ο λόγος για τον οποίο μία τέτοια καθιερωμένη τεχνική σαν το Listing ID αποτυγχάνει στην Airbnb, οφείλεται σε ορισμένες μοναδικές ιδιότητες της υποκείμενης αγοράς. Οι ενσωματώσεις χρειάζονται σημαντικές ποσότητες δεδομένων ανά στοιχείο για να συγκλίνουν σε λογικές τιμές. Όταν τα στοιχεία αυτά μπορούν να επαναλαμβάνονται δίχως περιορισμούς, όπως τα διαδικτυακά βίντεο ή οι λέξεις σε μια γλώσσα, δεν υπάρχει όριο στην ποσότητα αλληλεπίδρασης του χρήστη με ένα στοιχείο. Η απόκτηση μεγάλων ποσοτήτων για τα στοιχεία που μας ενδιαφέρουν είναι σχετικά εύκολη. Οι καταχωρίσεις από την άλλη πλευρά, υπόκεινται σε περιορισμούς από το φυσικό κόσμο. Ακόμα και η πιο δημοφιλής καταχώριση

μπορεί να κλείσει το πολύ 365 φορές σε έναν ολόκληρο χρόνο. Οι τυπικές κρατήσεις ανά καταχώριση είναι πολύ λιγότερες. Αυτός ο θεμελιώδης περιορισμός δημιουργεί δεδομένα που είναι πολύ αραιά σε επίπεδο καταχώρισης. Η υπερπροσφορά είναι μια άμεση επίπτωση αυτού του περιορισμού.

Αντιθέτως, όταν το μοντέλο Multi-task learning δοκιμάστηκε στο διαδίκτυο, αύξησε τις μακροχρόνιες προβολές με μεγάλη διαφορά, αλλά οι κρατήσεις παρέμειναν ουδέτερες. Ελέγχοντας χειροκίνητα τις καταχωρίσεις που είχαν υψηλή αναλογία μακροχρόνιων προβολών σε σύγκριση με τις κρατήσεις, διαπιστώθηκαν αρκετοί πιθανοί λόγοι οι οποίοι θα μπορούσαν να οδηγήσουν σε αυτό το κενό. Τέτοιες μακροχρόνιες προβολές θα μπορούσαν να οφείλονται σε καταχωρίσεις με υψηλές τιμές, σε καταχωρίσεις με μεγάλες περιγραφές που είναι δύσκολο να αναλυθούν, ή εξαιρετικά μοναδικές και μερικές φορές χιουμοριστικές καταχωρίσεις.

4.1.4. Συμπεράσματα

Οι συγκεκριμένοι ερευνητές από την Αμερική, ξεκίνησαν πιστεύοντας πως η βαθιά μάθηση θα ήταν μια σταγόνα στην αρχή της αντικατάστασης του μοντέλου GBDT και θα έδινε εκπληκτικά κέρδη. Πολλές από τις αρχικές συζητήσεις επικεντρώθηκαν στο να παραμείνουν όλα τα υπόλοιπα αμετάβλητα και να αντικατασταθεί μόνο το τρέχον μοντέλο με ένα νευρωνικό δίκτυο για να ελεγχθεί τι κέρδη θα μπορούσαν να υπάρξουν. Στην πραγματικότητα όμως, το μόνο που ανακαλύφθηκε, ήταν ότι δεν υπήρξαν κέρδη και το μόνο που υπήρξε στην αρχή ήταν η οπισθοδρόμηση στις μετρήσεις εκτός σύνδεσης. Με την πάροδο του χρόνου όμως διαπιστώθηκε πως η μετάβαση στη βαθιά μάθηση δεν είναι καθόλου μια απλή αντικατάσταση μοντέλου, αλλά μάλλον έχει να κάνει με την κλιμάκωση του συστήματος. Ως αποτέλεσμα, απαιτήθηκε επανεξέταση ολόκληρου του συστήματος που περιβάλλει το μοντέλο. Περιοριζόμενα σε μικρότερες κλίμακες, μοντέλα όπως το GBDT είναι αναμφισβήτητα ισάξια σε απόδοση και ευκολότερα στο χειρισμό και συνεχίζεται να χρησιμοποιείται για εστιασμένα προβλήματα μεσαίου μεγέθους.

4.2. Κατάταξη αναζήτησης και εξατομίκευση στην Airbnb.^[20]

4.2.1. Εισαγωγή

Η συγκεκριμένη έρευνα, πραγματοποιήθηκε από τον ερευνητή Mihajlo Grbovic τον Αύγουστο του 2017, και το θέμα της είναι η κατάταξη αναζήτησης και η εξατομίκευση στην Airbnb.

Κατά τη διάρκεια της τελευταίας δεκαετίας, οι αρχιτεκτονικές αναζήτησης, οι οποίες βασίζονταν συνήθως σε κλασσικές τεχνικές ανάκτησης πληροφοριών, έχουν δει μια αυξημένη εφαρμογή της μηχανικής μάθησης στα διάφορα στοιχεία της, ειδικά στην ανάκτηση αναζήτησης και στην κατάταξη αναζήτησης που συχνά έχουν δύσκολους στόχους ανάλογα με τον τύπο του περιεχομένου που αναζητείται. Ο κύριος λόγος πίσω από τη συγκεκριμένη τάση είναι η αύξηση αριθμού των δεδομένων αναζήτησης τα οποία μπορούν να συλλεχθούν και να αναλυθούν. Όσο περισσότερα δεδομένα γίνονται διαθέσιμα σχετικά με την αλληλεπίδραση των χρηστών με τα συστήματα ανάκτησης και κατάταξης αναζήτησης, τόσο περισσότερο μπορούν να βελτιωθούν, να βελτιστοποιηθούν και να προσαρμοστούν με τη χρήση της μηχανικής μάθησης. Επιπλέον, τα δεδομένα που συλλέγονται ανοίγουν δυνατότητες εξατομίκευσης της εμπειρίας για ένα συγκεκριμένο χρήστη βασισμένα στις προηγούμενες αναζητήσεις του ή τις προηγούμενες αναζητήσεις παρόμοιων χρηστών.

4.2.2. Χαρακτηριστικά και Μέθοδοι

Ο Mihajlo Grbovic, προσπάθησε να αναλύσει την περίπτωση της Airbnb, όπου υπάρχει σαφής ανάγκη για βελτιστοποίηση των αποτελεσμάτων αναζήτησης, τόσο για τους οικοδεσπότες των διαμερισμάτων, όσο και για τους επισκέπτες, πράγμα που σημαίνει ότι, δεδομένων των ερωτημάτων εισόδου με την τοποθεσία και τις ημερομηνίες ταξιδιού, πρέπει να καταταχθούν ψηλά οι καταχωρίσεις των οποίων η τοποθεσία, η τιμή, το στυλ, οι κριτικές κ.λπ. είναι ελκυστικές για τον επισκέπτη και ταυτόχρονα θα πρέπει να εντοπίζονται οι καταχωρίσεις για τις οποίες ο οικοδεσπότης πιθανότατα θα απέρριπτε τον επισκέπτη λόγω κακών κριτικών, αριθμού επισκεπτών, παιδιών, κατοικίδιων ζώων, καπνίσματος ή οποιουδήποτε άλλου παράγοντα. Οι καταχωρίσεις αυτές θα πρέπει να καταταχθούν χαμηλότερα. Για να επιτευχθεί αυτός ο τύπος κατάταξης, ο συγγραφέας της συγκεκριμένης έρευνας καταφεύγει σε τεχνικές εκπαίδευσης της κατάταξης.

Συγκεκριμένα, αξιοποιείται η ποιότητα των καταχωρίσεων, η συνάφεια της τοποθεσίας, οι κριτικές, ο χρόνος απόκρισης του οικοδεσπότη, καθώς και οι προτιμήσεις των επισκεπτών και των ιδιοκτητών των καταλυμάτων και το ιστορικό προηγούμενων κρατήσεων ως χαρακτηριστικά της κατάταξης. Το πρόβλημα διατυπώνεται ως παλινδρόμηση ανά ζεύγη με θετική χρησιμότητα για τις κρατήσεις και αρνητική χρησιμότητα για τις απορρίψεις, οι οποίες βελτιστοποιούνται χρησιμοποιώντας μια τροποποιημένη έκδοση του μοντέλου Lambda Rank,

το οποίο διεξάγει συγκρίσεις με ζεύγη καταχωρίσεων που βελτιστοποιούν από κοινού την κατάταξη, τόσο για τους επισκέπτες, όσο και για τους οικοδεσπότες. Δεδομένου ότι οι επισκέπτες συνήθως διεξάγουν πολλαπλές αναζητήσεις, κάνουν κλικ σε περισσότερες από μία καταχωρίσεις και επικοινωνούν με περισσότερους από έναν οικοδεσπότες κατά τη διάρκεια μίας συνεδρίας αναζήτησης, υπάρχει η δυνατότητα να χρησιμοποιηθούν αυτά τα σήματα εντός της συνεδρίας, δηλαδή τα μεγάλα κλικς, οι λίστες επιθυμιών, οι προσθήκες και οι επαφές, για την περαιτέρω εξατομίκευση των μελλοντικών αναζητήσεων και να εμφανίζεται στους χρήστες περιεχόμενο παρόμοιο με αυτό που έχουν ήδη δει. Για να μπορέσουν να υπολογιστούν οι ομοιότητες μεταξύ των καταχωρίσεων για σκοπούς εξατομίκευσης, χρησιμοποιούνται ιστορικές συνεδρίες αναζήτησης, δηλαδή τα αναγνωριστικά των καταχωρίσεων στις οποίες έκανε κλικ και επικοινωνήσε ο συγκεκριμένος χρήστης διατεταγμένα στο χρόνο, με σκοπό να εκπαιδευτεί η ενσωμάτωση των καταχωρίσεων. Αυτοί οι αλγόριθμοι ενσωμάτωσης είναι σε θέση να μαθαίνουν αναπαραστάσεις των καταχωρίσεων σε έναν συνεχή διανυσματικό χώρο χαμηλών διαστάσεων χρησιμοποιώντας το περιβάλλον άλλων καταχωρίσεων που χρησιμοποιούνται στην ίδια περίοδο λειτουργίας έτσι ώστε οι καταχωρίσεις που εμφανίστηκαν σε παρόμοιο πλαίσιο να είναι κοντά η μία στην άλλη στον συγκεκριμένο χώρο ενσωμάτωσης.

4.2.3. Αποτελέσματα

Διαπιστώθηκε ότι οι εκπαιδευμένες ενσωματώσεις καταχωρίσεων είναι σε θέση να καταγράψουν την τιμή, τον τύπο δωματίου, το στυλ, την τοποθεσία, την ποιότητα, την αξιολόγηση και πολλά άλλα, όλα αυτά σε ένα διάνυσμα χαμηλής διάστασης, π.χ. 32 κινητές μονάδες, και να κωδικοποιήσουν ομοιότητες πολύ καλύτερες από τις κλασσικές αναπαραστάσεις των καταχωρίσεων. Δεδομένων των ενσωματώσεων καταχώρισης, καθίσταται δυνατό να υπολογιστούν οι δυνατότητες εξατομίκευσης κατά τη διάρκεια της περιόδου λειτουργίας υπολογίζοντας ομοιότητες μεταξύ ενσωματώσεων καταχωρίσεων υποψηφίων που πρέπει να καταταχθούν και ενσωμάτωσης καταχωρίσεων που έχουν ήδη χρησιμοποιηθεί από τον επισκέπτη, δηλαδή καταχωρίσεις με κλικ, επαφή ή λίστα επιθυμιών. Κατά συνέπεια, το μοντέλο κατάταξης μας μαθαίνει ότι όσο υψηλότερη είναι αυτή η ομοιότητα τόσο υψηλότερα θα πρέπει να κατατάσσεται στη λίστα η υποψήφια καταχώριση.

4.2.4. Συμπεράσματα

Ο συγκεκριμένος ερευνητής, περιέγραψε μερικές από τις βασικές τεχνικές που χρησιμοποιούνται για την κατάταξη αναζήτησης στην Airbnb. Συγκεκριμένα, το πρόβλημα κατάταξης μίας καταχώρισης συνοψίζεται στην ιεράρχηση των καταχωρίσεων που απευθύνονται στον επισκέπτη, αλλά ταυτόχρονα υποβιβάζει καταχωρίσεις που πιθανότατα θα απέρριπταν τον επισκέπτη, κάτι που δεν επιλύεται εύκολα με τη χρήση ενός απλού γραμμικού

μοντέλου. Ο ερευνητής προσπάθησε να ρίξει φως στον τρόπο με τον οποίο βελτιστοποιούνται από κοινού οι δύο στόχοι, αξιοποιώντας την ποιότητα των καταχωρίσεων, τη συνάφεια της τοποθεσίας, τις κριτικές, το χρόνο απόκρισης του οικοδεσπότη καθώς και τον επισκέπτη, τις προτιμήσεις των επισκεπτών και του οικοδεσπότη και το ιστορικό προηγούμενων κρατήσεων. Ο συγγραφέας της συγκεκριμένης έρευνας, χρησιμοποίησε LambdaRank, η οποία είναι μια τεχνική όπου η κατάταξη μετατρέπεται σε πρόβλημα ταξινόμησης ή παλινδρόμησης ανά ζεύγη, ώστε να συγκριθούν οι καταχωρίσεις μεταξύ τους και να υπάρξει βελτίωση στην κατάταξη αναζήτησης.

4.3. Εκμάθηση μετρήσεων πρόθεσης κράτησης για την αναζήτηση της Airbnb.^[21]

4.3.1. Εισαγωγή

Η συγκεκριμένη έρευνα έλαβε χώρα στο Σαν Φρανσίσκο, την γενέτειρα πόλη της Airbnb, από τον ερευνητή Bradley Turnbull και δημοσιεύθηκε το Μάιο του 2019. Ο Bradley Turnbull, στην παρούσα έρευνα, έχει ως σκοπό να βοηθήσει στη βελτιστοποίηση του προβαλλόμενου περιεχομένου στην κατάταξη αναζήτησης της Airbnb, με σκοπό να βελτιώσει την εμπειρία των επισκεπτών. Δεδομένου όμως του υψηλού βαθμού κράτησης λιγότερων παραδοσιακών καταλυμάτων, οι χρήστες μπορούν να περάσουν πολλές ημέρες έως εβδομάδες ψάχνοντας και σαρώνοντας τη σελίδα περιγραφής των καταχωρίσεων καταλυμάτων προτού αποφασίσουν να κάνουν κάποια κράτηση. Επιπλέον, πολλές πληροφορίες σχετικά με μία καταχώριση είναι αδόμητες και μπορούν να βρεθούν από τον χρήστη αφού ελέγξει τις λεπτομέρειες της καταχώρισης στη σελίδα της καταχώρισης. Ως αποτέλεσμα, διαπιστώθηκε ότι οι παραδοσιακές μετρικές αναζήτησης δεν λειτουργούν καλά στο πλαίσιο της πλατφόρμας της Airbnb. Βασικές μετρικές των ενεργειών των χρηστών, όπως το ποσοστό των κλικ, ο αριθμός των καταχωρίσεων που είδαν, ή ο χρόνος παραμονής, δεν συσχετίζονται σταθερά με τις μεταγενέστερες επιχειρηματικές μετρήσεις. Για την αντιμετώπιση αυτών των ζητημάτων ο Bradley Turnbull χρησιμοποιεί μηχανική μάθηση για να απομονώσει σήματα πρόθεσης από πλούσια συμπεριφορικά δεδομένα. Αυτά τα σήματα έχουν βασικές εφαρμογές, όπως αναλυτικές γνώσεις, εισροές μοντελοποίησης κατάταξης και ταχύτητα πειραματισμού. Στη συγκεκριμένη έρευνα, περιγράφεται η ανάπτυξη ενός μοντέλου χρήστη που βασίζεται σε ένα μοντέλο πρόθεσης.

4.3.2. Χαρακτηριστικά και Μέθοδοι

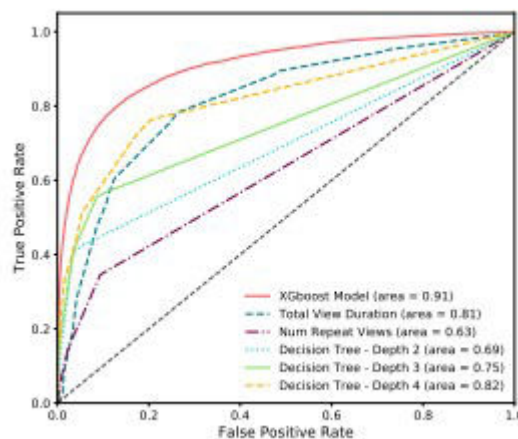
Ο ερευνητής της Airbnb Bradley Turnbull, ξεκίνησε εκπαιδώντας αρχεία καταγραφής χρηστών από μια σταθερή χρονική περίοδο και αξιολόγησε ένα ανεξάρτητο σύνολο δοκιμών από μια μελλοντική χρονική περίοδο. Κανένας χρήστης δεν αντιπροσωπεύεται και στα δύο σύνολα. Στη συνέχεια, πραγματοποίησε κάποιο βασικό φιλτράρισμα αφαιρώντας παραδείγματα από χρήστες με μία μόνο προβολή μιας καταχώρισης με διάρκεια μικρότερη από 5 δευτερόλεπτα. Προκειμένου να αποφύγουμε την υπερπροσφορά σε χρήστες με πολλές προβολές, σταθμίζουμε τα παραδείγματα κατάρτισης με το αμοιβαίο του συνολικού αριθμού παραδειγμάτων ανά χρήστη. Όλες οι δυνατότητες συνεχούς και καταμέτρησης μετασχηματίζονται. Έπειτα, ο ερευνητής χρησιμοποίησε τη βιβλιοθήκη XGBoost3 ώστε να εκπαιδεύσει το μοντέλο του, με αντικειμενική λειτουργία απώλειας αρχείων καταγραφής. Οι παράμετροι συντονισμού επιλέγονται αξιολογώντας τις επιδόσεις σε ένα ξεχωριστό σύνολο επικύρωσης.

4.3.3. Αποτελέσματα

Για τα αποτελέσματα, ο Bradley Turnbull χρησιμοποίησε AUC. Η AUC (Area Under the Curve) είναι το μέτρο της ικανότητας ενός ταξινομητή να διακρίνει μεταξύ των κλάσεων και χρησιμοποιείται ως περίληψη της καμπύλης ROC. Όσο υψηλότερη είναι η AUC, τόσο καλύτερη είναι η απόδοση του μοντέλου στη διάκριση μεταξύ των θετικών και αρνητικών κατηγοριών. Ο ερευνητής αξιολογεί την απόδοση AUC του μοντέλου ταξινόμησης XGboost σε σύγκριση με μία ποικιλία εναλλακτικών μοντέλων:

1. Συνολική διάρκεια προβολής - Βαθμολογείται το επίπεδο πρόθεσης της καταχώρισης από το άθροισμα της συνολικής διάρκειας προβολής της τρέχουσας προβολής και των προηγούμενων προβολών της καταχώρισης.
2. Αριθμός επαναλαμβανόμενων προβολών - Βαθμολογούνται οι προβολές με βάση τον συνολικό αριθμό των προηγούμενων προβολών της καταχώρισης από τον χρήστη.
3. Βάθος δέντρου αποφάσεων 2, 3 και 4 - Χρησιμοποιώντας τα ίδια χαρακτηριστικά όπως στο πλήρες μοντέλο XGboost, εκπαιδεύονται μερικά απλά δέντρα αποφάσεων διαφορετικού βάθους με το κριτήριο διαχωρισμού Gini.

Το Σχήμα 7 δείχνει την καμπύλη ROC και την απόδοση AUC για κάθε μοντέλο στο σύνολο δοκιμών. Το μοντέλο XGboost υπερτερεί σε απόδοση έναντι όλων των άλλων μοντέλων. Αυτό δεν αποτελεί έκπληξη δεδομένης της αυξημένης πολυπλοκότητας του του μοντέλου, αλλά η δραματική βελτίωση των επιδόσεων προσδίδει ένα ισχυρό επιχείρημα για τη συνέχιση της χρήσης του πιο σύνθετου μοντέλου αντί για κάτι απλούστερο.



Σχήμα 7. Αξιολόγηση της απόδοσης του μοντέλου.^[21]

4.3.4. Συμπεράσματα

Εν συντομία, ο ερευνητής παρουσιάζει μια αποτελεσματική προσέγγιση, χρησιμοποιώντας ένα μοντέλο ταξινόμησης με μηχανική μάθηση, για τη σύλληψη της πρόθεσης του χρήστη μέσω των σημάτων συμμετοχής του. Καθίσταται δυνατόν να μοντελοποιηθεί από κοινού, μέσω μίας συνθήκης πιθανότητας, η συνολική πρόθεση ενός χρήστη και ο τρόπος με τον οποίο αυτή η πρόθεση διασκορπίζεται σε μεμονωμένες καταχωρίσεις. Υπάρχει η δυνατότητα επίσης να επαναληφθεί η βελτίωση της απόδοσης πρόβλεψης του μοντέλου με τη μετάβαση σε νευρωνικά μοντέλα. Αυτό θα παρείχε μία πιο προηγμένη μέθοδο για την ενσωμάτωση των ιστορικών ενεργειών ενός χρήστη σε μια καταχώριση. Τέλος, διερευνούνται οι προβλέψεις της πρόθεσης προβολής ως ένα πιο αξιόπιστο σήμα ανατροφοδότησης στα μοντέλα κατάταξης αναζήτησης.

4.4. Εξατομίκευση σε πραγματικό χρόνο με τη χρήση ενσωματώσεων για την κατάταξη αναζήτησης στην Airbnb.^[22]

4.4.1. Εισαγωγή

Ο Mihajlo Grbovic, που τον είδαμε στην έρευνα για την Κατάταξη αναζήτησης και την εξατομίκευση στην Airbnb, στη συγκεκριμένη εργασία συνεργάζεται με έναν άλλο ερευνητή της Airbnb, τον Haibin Cheng, με σκοπό να βελτιώσουν την κατάταξη αναζήτησης της συγκεκριμένης πλατφόρμας.

Δεδομένης της μεγάλης διαφοράς στο περιεχόμενο που πρέπει να καταταχθεί, να εξατομικευτεί και να συστηθεί, κάθε αγορά έχει μια κάπως μοναδική πρόκληση. Αντίστοιχα, στην Airbnb, μια αγορά βραχυχρόνιας μίσθωσης, τα προβλήματα αναζήτησης και σύστασης είναι αρκετά μοναδικά, καθώς πρόκειται για μια αγορά δύο πλευρών στην οποία πρέπει να βελτιστοποιήσει κανείς τις προτιμήσεις του οικοδεσπότη και του φιλοξενούμενου, σε έναν κόσμο όπου ένας χρήστης σπάνια καταναλώνει το ίδιο αντικείμενο δύο φορές και μια καταχώριση μπορεί να δεχτεί μόνο έναν φιλοξενούμενο για μία συγκεκριμένη ημερομηνία.

Στη συγκεκριμένη έρευνα, οι Mihajlo Grbovic και Haibin Cheng, περιγράφουν τεχνικές ενσωμάτωσης καταχωρίσεων και χρηστών που ανέπτυξαν για σκοπούς εξατομίκευσης σε πραγματικό χρόνο στην κατάταξη αναζήτησης και στις συστάσεις παρόμοιων καταχωρίσεων. Τα μοντέλα ενσωμάτωσης προσαρμόστηκαν ειδικά για την αγορά της Airbnb και είναι σε θέση να καταγράφουν τα βραχυπρόθεσμα και μακροπρόθεσμα ενδιαφέροντα των επισκεπτών, παρέχοντας αποτελεσματικές συστάσεις για την καταχώριση καταλυμάτων. Οι δύο ερευνητές πραγματοποίησαν αυστηρές δοκιμές εκτός σύνδεσης των μοντέλων ενσωμάτωσης, ακολουθούμενες από επιτυχείς δοκιμές στο διαδίκτυο προτού τα αναπτύξουν πλήρως.

Στην περίπτωση της Airbnb, υπάρχει σαφής ανάγκη βελτιστοποίησης της αναζήτησης των αποτελεσμάτων τόσο για τους οικοδεσπότες όσο και για τους επισκέπτες, που σημαίνει ότι με δεδομένο ένα ερώτημα εισόδου με τοποθεσία και ημερομηνίες ταξιδιού πρέπει να καταταχθούν ψηλά οι καταχωρίσεις των οποίων η τοποθεσία, η τιμή, το στυλ, οι κριτικές κ.λπ. είναι ελκυστικές για τον επισκέπτη και, ταυτόχρονα, ταιριάζουν καλά με τις προτιμήσεις των οικοδεσποτών για τη διάρκεια του ταξιδιού του και τις ημέρες που επιθυμεί. Επιπλέον, πρέπει να εντοπιστούν οι καταχωρίσεις που πιθανόν να απορρίψουν τον επισκέπτη λόγω κακών κριτικών, κατοικίδιων ζώων, διάρκειας της διαμονής, του πλήθους των ατόμων ή οποιουδήποτε άλλου παράγοντα και να καταταχθούν χαμηλότερα.

4.4.2. Χαρακτηριστικά και Μέθοδοι

Συγκεκριμένα, οι δύο ερευνητές διατυπώνουν το πρόβλημα ως παλινδρόμηση ανά ζεύγη με θετικές παροχές για τις κρατήσεις και αρνητικές παροχές για τις απορρίψεις, τις οποίες

βελτιστοποιούν χρησιμοποιώντας μια τροποποιημένη έκδοση του μοντέλου Lambda Rank που βελτιστοποιεί από κοινού την κατάταξη και για τις δύο πλευρές της αγοράς. Δεδομένου ότι οι επισκέπτες συνήθως πραγματοποιούν πολλαπλές αναζητήσεις πριν από την κράτηση, δηλαδή κάνουν κλικ σε περισσότερες από μία καταχωρίσεις και επικοινωνούν με περισσότερους από έναν οικοδεσπότες κατά τη διάρκεια της συνεδρίας αναζήτησής τους, καθίσταται δυνατό να χρησιμοποιηθούν αυτά τα σήματα εντός της συνεδρίας, δηλ. κλικ, επαφές με οικοδεσπότες κ.λπ. για εξατομίκευση σε πραγματικό χρόνο, όπου ο στόχος είναι να δείξουμε στον επισκέπτη περισσότερες καταχωρίσεις παρόμοιες με αυτές που πιστεύουμε ότι του άρεσαν από την έναρξη της συνεδρίας αναζήτησης. Ταυτόχρονα δύναται να χρησιμοποιηθεί το αρνητικό σήμα, π.χ. παραλείψεις καταχωρίσεων υψηλής κατάταξης, για να δείξουμε στον επισκέπτη λιγότερες καταχωρίσεις παρόμοιες με αυτές που πιστεύουμε ότι δεν του άρεσαν.

Αρχικά, εκπαιδεύτηκαν κάποια Listing Embeddings. Στη συνέχεια, προβάλλεται το ιστορικό του Μοντέλου Κατάταξης Αναζήτησης και περιγράφεται πώς χρησιμοποιήθηκαν οι Listing Embeddings και οι Listing Type & User Type Embeddings για την υλοποίηση χαρακτηριστικών για την Εξατομίκευση σε Πραγματικό Χρόνο στην Αναζήτηση.

4.4.3. Αποτελέσματα

Οι συγγραφείς της συγκεκριμένης έρευνας, πραγματοποίησαν πειράματα offline και online (A/B test). Πρώτον, συνέκριναν δύο μοντέλα κατάταξης αναζήτησης που εκπαιδεύτηκαν στα ίδια δεδομένα με και χωρίς χαρακτηριστικά ενσωμάτωσης. Στον Πίνακα 8 συνοψίζονται τα αποτελέσματα του DCU (Discounted Cumulative Utility) ανά κάθε χρησιμότητα (εντύπωση, κλικ, απόρριψη και κράτηση) και του συνολικού NDCU (Normalized Discounted Cumulative Utility). Παρατηρείται ότι η προσθήκη χαρακτηριστικών ενσωμάτωσης οδήγησε σε αύξηση της NDCU κατά 2,27%, όπου η κράτηση DCU αυξήθηκε κατά 2,58%, πράγμα που σημαίνει ότι οι καταχωρίσεις με κράτηση κατατάχθηκαν υψηλότερα στο σύνολο των κρατήσεων, χωρίς να επηρεάζονται οι απορρίψεις (DCU -0,4 παρέμεινε σταθερό), που σημαίνει ότι οι καταχωρίσεις που απορρίφθηκαν δεν κατατάχθηκαν υψηλότερα από ό,τι στο μοντέλο χωρίς ενσωματωμένα χαρακτηριστικά.

Metrics	Percentage Lift
DCU -0.4 (rejections)	+0.31%
DCU 0.01 (clicks)	+1.48%
DCU 0.25 (contacts)	+1.95%
DCU 1 (bookings)	+2.58%
NDCU	+2.27%

Πίνακας 8. Αποτελέσματα offline αξιολόγησης.^[22]

Οι παρατηρήσεις από τον Πίνακα 8, καθώς και το γεγονός ότι η ενσωμάτωση χαρακτηριστικών κατατάσσονται ψηλά στις εισαγωγές χαρακτηριστικών GBDT και επίσης η διαπίστωση ότι η συμπεριφορά των χαρακτηριστικών ταιριάζει με αυτό που περίμεναν οι ερευνητές, ήταν αρκετά για να πάρουν την απόφαση να προχωρήσουν σε ένα online πείραμα. Στο online πείραμα, παρατήρησαν ένα στατιστικά σημαντικό κέρδος κρατήσεων και τα χαρακτηριστικά ενσωμάτωσης ξεκίνησαν την παραγωγή. Αρκετούς μήνες αργότερα, οι ερευνητές διενέργησαν μια αντίστροφη δοκιμή στην οποία επιχειρήσαν να αφαιρέσουν τα χαρακτηριστικά ενσωμάτωσης, το οποίο είχε ως αποτέλεσμα αρνητικές κρατήσεις. Αυτό ήταν άλλη μια ένδειξη ότι τα real-time χαρακτηριστικά ενσωμάτωσης είναι αποτελεσματικά.

4.4.4. Συμπεράσματα

Στη συγκεκριμένη έρευνα, ο Mihajlo Grbovic και ο Haibin Cheng, πρότειναν μια νέα μέθοδο για εξατομίκευση σε πραγματικό χρόνο στην κατάταξη αναζήτησης στην Airbnb. Η μέθοδος μαθαίνει χαμηλής διάστασης αναπαραστάσεις των καταχωρίσεων κατοικιών και των χρηστών με βάση τη συνύπαρξη στις συνεδρίες κλικ και κρατήσεων των χρηστών. Για την καλύτερη αξιοποίηση των διαθέσιμων πλαισίων αναζήτησης, ενσωμάτωσαν έννοιες όπως το παγκόσμιο πλαίσιο και τα ρητά αρνητικά σήματα στη διαδικασία εκπαίδευσης. Αξιολόγησαν την προτεινόμενη μέθοδο σε παρόμοιες προτάσεις καταχωρίσεων και στην κατάταξη αναζήτησης. Μετά από επιτυχή δοκιμή σε ζωντανή κίνηση αναζήτησης αναπτύχθηκαν και οι δύο εφαρμογές ενσωμάτωσης επιτυχώς.

4.5. Γενικά συμπεράσματα και μελλοντικές κατευθύνσεις

Στην πρώτη έρευνα του συγκεκριμένου κεφαλαίου, η οποία είχε να κάνει με την εφαρμογή βαθιάς μάθησης στην αναζήτηση της Airbnb, οι ερευνητές από την Αμερική, ξεκίνησαν πιστεύοντας πως η βαθιά μάθηση θα ήταν μια σταγόνα στην αρχή της αντικατάστασης του μοντέλου GBDT και θα απέφερε εκπληκτικά κέρδη. Διαπίστωσαν όμως ότι δεν υπήρξαν κέρδη και συμπέραναν ότι μοντέλα όπως το GBDT είναι αναμφισβήτητα ισάξια σε απόδοση και ευκολότερα στο χειρισμό.

Στην επόμενη έρευνα που αναλύθηκε, ο συγκεκριμένος ερευνητής περιέγραψε μερικές από τις βασικές τεχνικές που χρησιμοποιούνται για την κατάταξη αναζήτησης στην Airbnb. Για να γίνει εφικτή η προσπάθειά του, χρησιμοποίησε LambdaRank, η οποία είναι μία τεχνική όπου η κατάταξη μετατρέπεται σε πρόβλημα ταξινόμησης ή παλινδρόμησης ανά ζεύγη, ώστε να συγκριθούν οι καταχωρίσεις μεταξύ τους και να υπάρξει βελτίωση στην κατάταξη αναζήτησης.

Από μία άλλη οπτική, ο ερευνητής Bradley Turnbull, χρησιμοποίησε XGBoost3 δοκιμάζοντάς το μαζί με πολλά διαφορετικά μοντέλα και κατέληξε στο συμπέρασμα ότι το συγκεκριμένο

μοντέλο υπερτερεί έναντι των υπολοίπων για την εκμάθηση προθέσεων κράτησης στην αναζήτηση της Airbnb.

Στην τελευταία έρευνα του συγκεκριμένου κεφαλαίου, οι ερευνητές Mihajlo Grbovic και Haibin Cheng, πρότειναν μια νέα μέθοδο για εξατομίκευση σε πραγματικό χρόνο στην κατάταξη αναζήτησης στην Airbnb. Η μέθοδος μαθαίνει χαμηλής διάστασης αναπαραστάσεις των καταχωρίσεων κατοικιών και των χρηστών με βάση τη συνύπαρξη στις συνεδρίες κλικ και κρατήσεων των χρηστών.

Τέλος, οι ερευνητές στο μέλλον θα μπορούσαν να διερευνήσουν εναλλακτικές μεθόδους για τη βελτίωση της κατάταξης αναζήτησης στην Airbnb, όπως και να βελτιώσουν τις αποδόσεις πρόβλεψης των μοντέλων με τη μετάβαση σε μοντέλα ακολουθίας νευρωνικών δικτύων. Η συγκεκριμένη ενέργεια θα μπορούσε να παρέχει μια πιο προηγμένη μέθοδο για την ενσωμάτωση των ιστορικών ενεργειών ενός χρήστη σε μια καταχώριση και έτσι να μη χρησιμοποιούνται απλά στατιστικά στοιχεία όπως γίνεται σήμερα.

5. ΔΥΝΑΜΙΚΗ ΤΙΜΟΛΟΓΗΣΗ

5.1. Προσαρμοσμένο μοντέλο παλινδρόμησης για δυναμική τιμολόγηση στην Airbnb.^[23]

5.1.1. Εισαγωγή

Οι συγγραφείς της συγκεκριμένης έρευνας, που δημοσιεύθηκε τον Αύγουστο του 2018, περιγράφουν το μοντέλο στρατηγικής τιμολόγησης που εφαρμόζεται στην Airbnb. Ο στόχος της βελτιστοποίησης των τιμών, είναι να βοηθήσει τους οικοδεσπότες της Airbnb ώστε να ορίσουν τη βέλτιστη τιμή για τις καταχωρίσεις τους. Σε αντίθεση με τα συμβατικά προβλήματα τιμολόγησης, όπου οι στρατηγικές τιμολόγησης εφαρμόζονται σε μεγάλη ποσότητα πανομοιότυπων προϊόντων, δεν υπάρχουν «πανομοιότυπα» προϊόντα στην Airbnb, διότι κάθε καταχώριση στην πλατφόρμα προσφέρει μοναδικές αξίες και εμπειρίες στους επισκέπτες. Επειδή η Airbnb είναι ιδιαίτερη στον τομέα των καταχωρίσεων, καθίσταται αρκετά δύσκολη η εκτίμηση μιας ακριβούς καμπύλης ζήτησης που απαιτείται για την εφαρμογή συμβατικών στρατηγικών τιμολόγησης μεγιστοποίησης εσόδων.

Το σύστημα τιμολόγησης το οποίο ανέπτυξαν οι ερευνητές, αποτελείται από τρία στοιχεία. Πρώτον, ένα δυαδικό μοντέλο ταξινόμησης το οποίο προβλέπει την πιθανότητα κράτησης κάθε νύχτας καταχώρισης. Δεύτερον, ένα μοντέλο παλινδρόμησης το οποίο προβλέπει τη βέλτιστη τιμή για κάθε νύχτα καταχώρισης, στην οποία χρησιμοποιείται μια προσαρμοσμένη λειτουργία απώλειας για να καθοδηγήσει τη μάθηση.

Σε αυτή την έρευνα, εστιάζεται η περιγραφή του μοντέλου παλινδρόμησης στο δεύτερο στάδιο του συστήματος τιμολόγησης. Περιγράφεται επίσης ένα νέο σύνολο μετρήσεων για αξιολόγηση εκτός σύνδεσης. Η προτεινόμενη στρατηγική τιμολόγησης έχει αναπτυχθεί στην παραγωγή για να τροφοδοτήσει το εργαλείο Price Tips και Smart Pricing στην Airbnb. Τα αποτελέσματα των online δοκιμών A/B αποδεικνύουν την αποτελεσματικότητα του προτεινόμενου μοντέλου στρατηγικής.

5.1.2. Χαρακτηριστικά και Μέθοδοι

Το σύστημα τιμολόγησης (που παρουσιάζεται στο Σχήμα 8), αποτελείται από τρία κύρια στοιχεία. Πρώτον, ένα μοντέλο δυαδικής ταξινόμησης προβλέπει την πιθανότητα κράτησης για κάθε διανυκτέρευση. Δεύτερον, ένα μοντέλο παλινδρόμησης ανά καταχώριση, προβλέπει τη βέλτιστη τιμή για κάθε νύχτα, στην οποία χρησιμοποιείται μια προσαρμοσμένη συνάρτηση απωλειών για την καθοδήγηση της εκπαίδευσης του μοντέλου. Η προβλεπόμενη πιθανότητα κράτησης από το πρώτο μοντέλο χρησιμοποιείται ως χαρακτηριστικό εισόδου για το μοντέλο παλινδρόμησης. Τέλος, εφαρμόζεται πρόσθετη λογική εξατομίκευσης που ενσωματώνει και

στόχους φιλοξενίας, ειδικές εκδηλώσεις κ.λπ. πάνω από το αποτέλεσμα του δεύτερου μοντέλου ώστε να δημιουργηθούν οι τελικές προτάσεις τιμών.



Σχήμα 8. Επισκόπηση του συστήματος τιμολόγησης.^[23]

Προκειμένου να εκτιμηθεί η καμπύλη ζήτησης για κάθε νύχτα καταχώρισης, οι ερευνητές κατασκεύασαν ένα μοντέλο πιθανότητας κράτησης. Ο στόχος της εκπαίδευσης του μοντέλου είναι να προβλεφθεί αν μια διαθέσιμη μελλοντική νύχτα μιας συγκεκριμένης καταχώρισης θα είναι κλεισμένη από την ημέρα που γίνεται η πρόβλεψη, το οποίο είναι ένα τυπικό πρόβλημα δυαδικής ταξινόμησης. Ακολουθούν παραδείγματα χαρακτηριστικών που χρησιμοποιήθηκαν στο συγκεκριμένο μοντέλο.

- Χαρακτηριστικά καταχώρισης: τιμή της καταχώρισης ανά διανυκτέρευση, τύπος του δωματίου, χωρητικότητα ατόμων, αριθμός υπνοδωματίων / μπάνιων, ανέσεις, τοποθεσίες, κριτικές, ιστορικό ποσοστό πληρότητας, ενεργοποιημένη άμεση κράτηση, κ.λπ.
- Χρονικά χαρακτηριστικά: εποχικότητα (η ημέρα του έτους, η ημέρα της εβδομάδας, κ.λπ.), η ημερολογιακή διαθεσιμότητα (π.χ. το κενό μεταξύ του check in και του check out), απόσταση μεταξύ ds και ds_night, κ.λπ.
- Δυναμική προσφοράς και ζήτησης: αριθμός των διαθέσιμων καταχωρίσεων στη γειτονιά, προβολές καταχωρίσεων, ποσοστά αναζητήσεων / επαφών, κ.λπ.

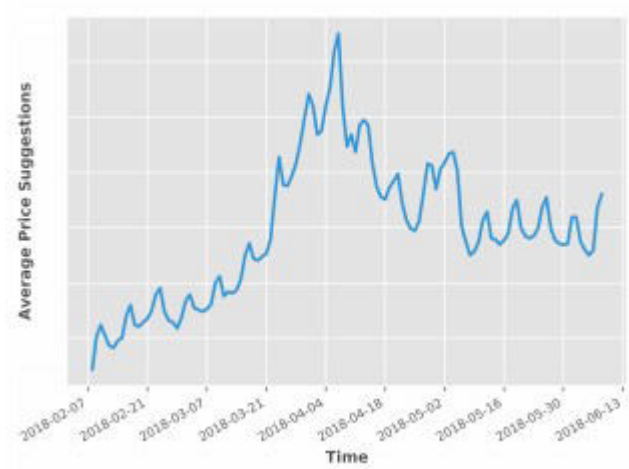
Επίσης, οι ερευνητές χρησιμοποιούν το Spark για τον παραλληλισμό της εκπαίδευσης. Τα μοντέλα εκπαιδεύονται και ενημερώνονται σε καθημερινή βάση. Χρησιμοποιείται επίσης η μέθοδος Stochastic Gradient Descent (SGD) για τη βελτιστοποίηση της συνάρτησης απώλειας.

5.1.3. Αποτελέσματα

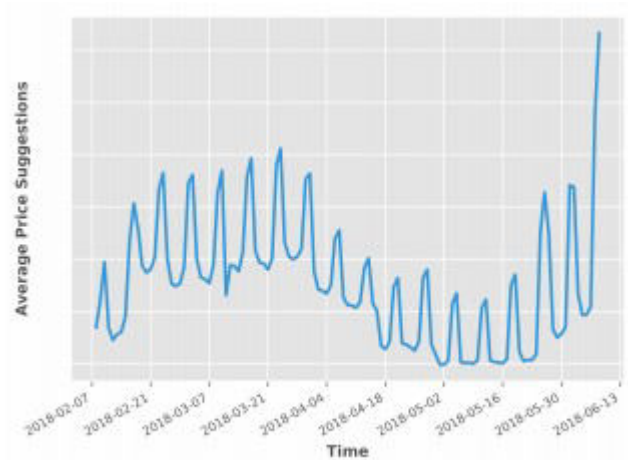
Οι ερευνητές χρησιμοποίησαν ένα Gradient Boosting Machine (GBM), με σκοπό να προβλέψουν τις πιθανότητες κράτησης. Αντί να εκπαιδεύσουν ένα παγκόσμιο μοντέλο GBM με

σταθερό ρυθμό δειγματοληπτικών δεδομένων εκπαίδευσης, διαπίστωσαν ότι καλύτερη απόδοση όσον αφορά τη συνολική AUC, μπορεί να επιτευχθεί με την εκπαίδευση ενός ξεχωριστού μοντέλου GBM για κάθε αγορά με προσαρμοστικό ρυθμό δειγματοληπτικών δεδομένων εκπαίδευσης. Οι αγορές με υψηλή πυκνότητα καταχωρίσεων επωφελούνται περισσότερο από τα μοντέλα που βασίζονται στην τοποθεσία, τα οποία δοκιμάζονται με ρυθμό υψηλότερο από τον παγκόσμιο σταθερό ρυθμό δειγματοληψίας.

Πέρα από τις αξιολογήσεις σε σύνδεση και εκτός σύνδεσης, γίνανε και προτάσεις τιμών επιτόπιου ελέγχου, ώστε να ελεγχθεί αν οι προτάσεις για τις τιμές ανταποκρίνονται στις αλλαγές της ζήτησης. Συγκεκριμένα, οι ερευνητές επέλεξαν δειγματοληπτικά τις τελικές προτάσεις τιμών που δημιουργήθηκαν στις 08/02/2018 για 120 μελλοντικές διανυκτερεύσεις. Τα Σχήματα 9 και 10 δείχνουν πώς οι μέσες προτάσεις τιμών μεταβάλλονται με την πάροδο του χρόνου για το Tokyo και την Tahoe. Και για τις δύο αγορές, βλέπουμε ότι υπάρχει ισχυρό εβδομαδιαίο μοτίβο. Το μοντέλο τείνει να προτείνει υψηλότερες τιμές το Σαββατοκύριακο. Για το Τόκιο, παρατηρείται μια έντονη αιχμή από τα τέλη Μαρτίου έως τις αρχές Απριλίου. Από τα δύο αυτά παραδείγματα, βλέπουμε ότι το μοντέλο μπορεί να αποτυπώσει τη δυναμική της αγοράς εγκαίρως.



Σχήμα 9. Tokyo.^[23]



Σχήμα 10. Tahoe, CA.^[23]

5.1.4. Συμπεράσματα

Στη συγκεκριμένη εργασία, οι ερευνητές περιέγραψαν τις προκλήσεις του προβλήματος της δυναμικής τιμολόγησης στην Airbnb και παρουσίασαν ένα μοντέλο στρατηγικής τιμολόγησης που αναπτύχθηκε με σκοπό να βοηθήσει τους οικοδεσπότες της Airbnb να καθορίζουν τις τιμές τους αποτελεσματικότερα. Τα αποτελέσματα αξιολόγησης εκτός σύνδεσης και σε σύνδεση δείχνουν ότι το προτεινόμενο μοντέλο στρατηγικής αποδίδει σημαντικά καλύτερα από ένα άμεσο στρατηγικό μοντέλο τιμολόγησης max-rev. Οι ερευνητές συνεχίζουν να εργάζονται για τη βελτίωση της εκτίμησης της καμπύλης ζήτησης. Με μια πιο ακριβή καμπύλη ζήτησης, υπάρχει η δυνατότητα να επανεξεταστεί η άμεση στρατηγική μεγιστοποίησης εσόδων στο μέλλον.

5.2. Διερεύνηση της χρήσης τεχνητής νοημοσύνης στη μεγιστοποίηση των τιμών στον τομέα του τουρισμού: η εφαρμογή της στην περίπτωση της Airbnb στην Κοινότητα της Βαλένθια.^[24]

5.2.1. Εισαγωγή

Η συγκεκριμένη έρευνα, δημοσιεύθηκε το 2018 από τέσσερις Ισπανούς ερευνητές οι οποίοι προσπαθούν να διερευνήσουν τη χρήση της τεχνητής νοημοσύνης και συγκεκριμένα της μηχανικής μάθησης πάνω στη μεγιστοποίηση των τιμών στον τομέα του τουρισμού για την περίπτωση της Airbnb στην Κοινότητα της Βαλένθια.

Η χρήση της μηχανικής μάθησης γίνεται όλο και συχνότερη στην αναζήτηση ανταγωνιστικότητας από τις επιχειρήσεις. Στη συγκεκριμένη έρευνα, οι ερευνητές εξετάζουν μέσα από τη βιβλιογραφία, το πώς η τεχνητή νοημοσύνη μπορεί να βοηθήσει τις εταιρίες να βελτιώσουν τις γνώσεις τους σχετικά με τους χρήστες, να βελτιστοποιήσουν τις τιμές τους ή να καθοδηγήσουν τους αγοραστές σε συγκεκριμένες επιλογές. Για να επιβεβαιωθεί ότι η εφαρμογή μοντέλων τεχνητής νοημοσύνης επιτρέπει στις εταιρείες να αποκτήσουν καλύτερες διαδικασίες βελτιστοποίησης των τιμών από ότι με άλλα παραδοσιακά μοντέλα, οι ερευνητές μελέτησαν περισσότερα από 10.000 ακίνητα Airbnb στις τρεις κύριες πόλεις της κοινότητας της Βαλένθια (Βαλένθια, Αλικάντε, Καστεγιόν), σημειώνοντας ότι η διαδικασία εκτίμησης με τη χρήση νευρωνικών δικτύων προσφέρει σημαντικά πιο ικανοποιητικά αποτελέσματα από τη χρήση ηδονικών μοντέλων.

5.2.2. Χαρακτηριστικά και Μέθοδοι

Για τη συγκεκριμένη έρευνα, οι τέσσερις Ισπανοί ερευνητές χρησιμοποίησαν τη μέθοδο της ηδονικής τιμολόγησης και τη μέθοδο των τεχνητών νευρωνικών δικτύων. Η επιλογή και η χρήση των συγκεκριμένων μεθόδων αναλύεται παρακάτω:

1. Ηδονική μέθοδος τιμολόγησης

Η ηδονική τιμολόγηση είναι ένα μοντέλο που προσδιορίζει τους παράγοντες της τιμής σύμφωνα με την παραδοχή ότι η τιμή καθορίζεται τόσο από τα εσωτερικά χαρακτηριστικά του αγαθού που πωλείται όσο και από εξωτερικούς παράγοντες που το επηρεάζουν. Όταν προσδιορίζεται η αξία ενός ακινήτου, είτε πρόκειται για ενοικίαση είτε για πώληση, το μοντέλο ηδονικής τιμολόγησης αποτελεί εδώ και δεκαετίες μια από τις πιο ευρέως χρησιμοποιούμενες μεθόδους, με την αντίληψη ότι η τιμολόγηση βασίζεται σε μια σειρά από στοιχεία που συμβάλλουν με μια ορισμένη αξία στο τελικό αγαθό (Rigall-I-Torrent, R. και συνεργάτες, 2011). Εκ πρώτης όψεως, είναι αδύνατο να προσδιοριστούν όλα τα χαρακτηριστικά που συνθέτουν ένα προϊόν εξ ολοκλήρου, καθώς δεν πρέπει να ληφθούν υπόψη μόνο φυσικοί ή μετρήσιμοι παράγοντες, αλλά και ψυχολογικοί ή χρονικοί. Η

μεθοδολογία των ηδονικών τιμών επιτρέπει την εκτίμηση της αξίας που συνεισφέρει κάθε ένα από τα χαρακτηριστικά (φυσικά ή άλλα) σε ένα ακίνητο και να γίνουν προβλέψεις σχετικά με τη συμπεριφορά των υπόλοιπων ιδιοτήτων όταν κάποιο από αυτά τα στοιχεία μεταβάλλεται. Η μεθοδολογία των ηδονικών τιμών είναι ιδιαίτερα χρήσιμη για τη μέτρηση της επίδρασης συγκεκριμένων στοιχείων στην αξία των αγαθών ή υπηρεσιών μέσω μιας ανάλυσης πολλαπλής παλινδρόμησης. Με αυτόν τον τρόπο, η συνάρτηση της τιμής (P) που λαμβάνει κάθε ένα από τα χαρακτηριστικά ή στοιχεία μπορεί να εκφραστεί ως εξής:

$$P_i = \beta_0 + \beta_1\chi_{1i} + \beta_2\chi_{2i} + \dots + \beta_j\chi_{ji} + e_i ,$$

όπου P_i είναι η τιμή του αγαθού i , και κάθε ένα από τα β_j είναι τα χαρακτηριστικά που ορίζονται με τους αντίστοιχους συντελεστές παλινδρόμησης. Τέλος, το «e» αντιπροσωπεύει το περιθώριο σφάλματος.

2. Τεχνητά νευρωνικά δίκτυα

Το μοντέλο τεχνητών νευρωνικών δικτύων (ΤΝΔ), από την άλλη πλευρά, γεννήθηκε ως μέρος της έρευνας στον τομέα της τεχνητής νοημοσύνης, η οποία προσπαθούσε να διαμορφώσει τη δομή του ανθρώπινου εγκεφάλου και να αναπαράγει την ικανότητα των βιολογικών νευρικών συστημάτων να μελετούν και να διορθώνουν λάθη. Κατά κάποιο τρόπο, τα τεχνητά νευρωνικά δίκτυα είναι ένα μοντέλο τεχνητής νοημοσύνης που έχει σχεδιαστεί για να αναπαράγει τις μαθησιακές διαδικασίες του ανθρώπινου εγκεφάλου, να γενικεύει τα αποτελέσματα και να ανταποκρίνεται με τον πιο λογικό δυνατό τρόπο σε περιπτώσεις απουσίας ή άγνοιας δεδομένων (Shaw, 1992). Τα νευρωνικά δίκτυα καθιστούν δυνατή τη δημιουργία και αναγνώριση μοτίβων και παρέχουν την ικανότητα να επεξεργάζονται πληροφορίες που περιλαμβάνουν τυχαίες συνιστώσες, με την δυνατότητα εφαρμογής σε παραδοσιακές στατιστικές τεχνικές (Otero and Trujillo, 1993).

Σε αντίθεση με το ηδονικό μοντέλο τιμολόγησης, η μεθοδολογία των τεχνητών νευρωνικών δικτύων επιτρέπει την εντοπισμό μη γραμμικών σχέσεων μεταξύ μεταβλητών και, ως εκ τούτου, μπορεί δυνητικά να παρέχει καλύτερη προγνωστική ικανότητα από την πολυ-αμετάβλητη ανάλυση, σύμφωνα με τον Wilson και τους συνεργάτες του (2002).

Σκοπός της παρούσας έρευνας, είναι να συγκριθούν τα δύο αυτά μοντέλα μεταξύ τους, προκειμένου να προσδιοριστεί ποιο από αυτά είναι πιο αξιόπιστο όσον αφορά την πρόβλεψη της αξίας των ενοικιαζόμενων ακινήτων στην Airbnb. Αυτή η διαδικασία σύγκρισης έχει ήδη πραγματοποιηθεί από ορισμένους προηγούμενους συγγραφείς, όπως ο Selim (2011) στην τουρκική αγορά για την αξία των ακινήτων το 2004, στην οποία συνέκρινε τα αποτελέσματα πρόβλεψης που επιτεύχθηκαν μεταξύ της ηδονικής παλινδρόμησης και των μοντέλων τεχνητού δικτύου και κατέδειξε τη δυνατότητα των τελευταίων να είναι μια καλύτερη εναλλακτική λύση

από την πρώτη στην πρόβλεψη της τιμής της κατοικίας. Παρόμοιο αποτέλεσμα επιτεύχθηκε στη μελέτη που πραγματοποιήθηκε για 200 ακίνητα στη Νέα Ζηλανδία από τον Limsombunchai και τους συνεργάτες του (2004), στην οποία η επίδραση παραγόντων όπως το μέγεθος του ακινήτου, ο αριθμός των υπνοδωματίων ή η θέση του, μεταξύ άλλων, αναλύθηκε και πάλι από την ηδονική παλινδρόμηση και τα νευρωνικά δίκτυα. Οι Peterson και Flanagan (2009) τόνισαν για άλλη μια φορά ότι τα τεχνητά νευρωνικά δίκτυα προκάλεσαν λιγότερα σφάλματα στην πρόβλεψη των τιμών της αξίας των κατοικιών για ένα δείγμα περισσότερων από 45.000 σπιτιών στη Βόρεια Καρολίνα. Ωστόσο, δεν έχει εφαρμοστεί μέχρι σήμερα σε σύγκριση μεταξύ των τουριστικών ενοικιαζόμενων καταλυμάτων, όπως συμβαίνει με την Airbnb.

5.2.3. Αποτελέσματα

Στόχος της συγκεκριμένης έρευνας είναι η εκτίμηση της αξίας ενοικίασης στην Airbnb μέσω μιας σειράς καθορισμένων χαρακτηριστικών για το ίδιο το διαμέρισμα, τα οποία σχετίζονται με την πόλη στην οποία βρίσκεται το ακίνητο. Το δείγμα περιλαμβάνει τα διαμερίσματα που προσφέρονται στην Airbnb στις τρεις επαρχιακές πρωτεύουσες της Κοινότητας της Βαλένθια (Αλικάντε, Βαλένθια και Καστεγιόν) κατά τους μήνες Μάιο του 2016 έως Μάιο του 2017. Η Κοινότητα της Βαλένθια είναι ένας από τους σημαντικότερους τουριστικούς προορισμούς στην Ευρώπη, με περισσότερους από εννέα εκατομμύρια τουρίστες ετησίως. Ωστόσο, οι τρεις πρωτεύουσες της περιοχής συνδυάζουν τις συνθήκες του ήλιου και της θάλασσας με μια πλήρη παροχή υπηρεσιών (όπως πανεπιστήμια ή δημόσια και ιδιωτικά νοσοκομεία), καθώς και πολιτιστικές δραστηριότητες, ιστορικά μνημεία και διάσημα εστιατόρια. Αυτό συνεπάγεται σημαντικές διαφορές στη σύνθεση της Airbnb σε σχέση με τους προορισμούς που εξαρτώνται πλήρως από τον καλοκαιρινό τουρισμό: Οι τρεις πρωτεύουσες έχουν χαμηλότερη εποχικότητα της τουριστικής ζήτησης και μικρότερη αναλογία δεύτερων κατοικιών από τον μέσο όρο της περιφέρειας της Βαλένθια (36%). Και στις τρεις πόλεις που μελετήθηκαν (ειδικά στο Αλικάντε και στη Βαλένθια) υπήρξε μεγάλη αύξηση των ακινήτων που δημοσιεύονται στην Airbnb από το 2014, συνδυάζοντας τόσο διαμερίσματα όσο και δωμάτια (πλήρη και κοινόχρηστα). Ωστόσο, υπάρχει μεγάλος αριθμός αυτών των ακινήτων που δεν έχουν ενοικιαστεί ποτέ. Αυτή είναι μια φυσιολογική κατάσταση στη συνεργατική αγοράς συνεργατικής οικονομίας όπου υπάρχει υπερπροσφορά αγαθών (διαμερίσματα) και πλήρης πληροφόρηση.

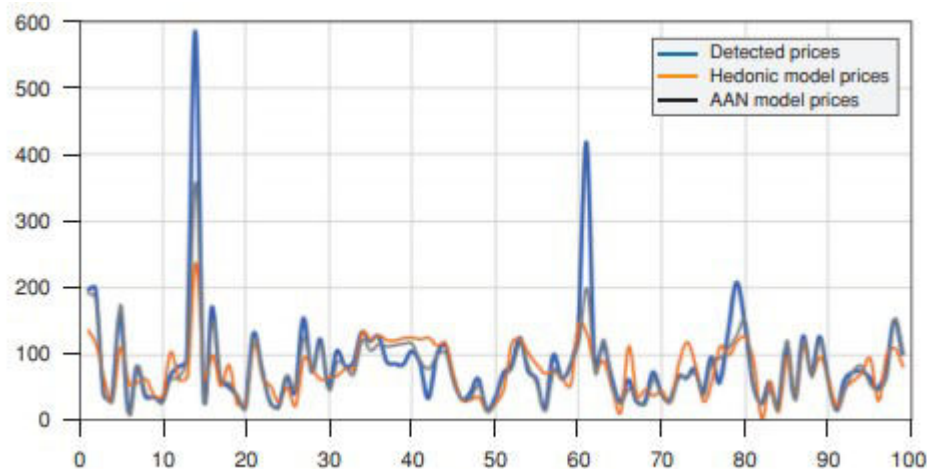
Από όλα τα σπίτια που ήταν διαθέσιμα από τον Μάιο του 2016 έως τον Μάιο του 2017 στην Κοινότητα της Βαλένθια, ελήφθησαν υπόψη μόνο εκείνα που ενοικιάστηκαν για τουλάχιστον μία ημέρα, αφήνοντας εκτός τις κατοικίες που δεν ενοικιάστηκαν καθόλου. Επομένως, η εξαρτημένη μεταβλητή της έρευνας είναι η μέση τιμή ενοικίασης, λαμβάνοντας υπόψη μόνο εκείνα τα ακίνητα που ενοικιάστηκαν για τουλάχιστον μία ημέρα εντός της προαναφερθείσας περιόδου.

Όσον αφορά την εκτίμηση των μοντέλων, ελήφθησαν υπόψη 8.096 ακίνητα στην πόλη της Βαλένθια, 2.988 ακίνητα στην πόλη του Αλικάντε και 299 ακίνητα στο Καστεγιόν, με σημαντικές διαφορές το καθένα. Στην πόλη του Καστεγιόν, για παράδειγμα, οι ενοικιάσεις μονών ή κοινόχρηστων δωματίων κυριαρχούν έναντι των ολόκληρων ακινήτων (56% έναντι 44%), ενώ στη Βαλένθια, τα ολόκληρα ακίνητα αντιπροσωπεύουν το 60% της προσφοράς και στο Αλικάντε το 70%. Αν και οι συνολικοί αυτοί αριθμοί μπορεί να φαίνονται πολύ υψηλοί, πρέπει να θυμόμαστε ότι η συνεργατική οικονομία βασίζεται στη βασική υπόθεση της "πλεονάζουσας προσφοράς" (MorenoIzquierdo και συνεργάτες, 2015) και στην προκειμένη περίπτωση μπορεί να παρατηρηθεί ξεκάθαρα εξετάζοντας το ποσοστό πληρότητας. Στην περίπτωση του Καστεγιόν, το μέσο ποσοστό πληρότητας των ακινήτων είναι 40% και 30% στην περίπτωση των μονόκλινων και κοινόχρηστων δωματίων. Στην περίπτωση του Αλικάντε, βλέπουμε ένα ποσοστό πληρότητας 43% για τα ολόκληρα διαμερίσματα και 35% για τα δωμάτια. Όσον αφορά τη Βαλένθια, υπάρχει ποσοστό πληρότητας 48% στην περίπτωση των ολόκληρων διαμερισμάτων, και 40% στην περίπτωση της ενοικίασης δωματίων.

Κατά τη διαδικασία ανάπτυξης του τεχνητού νευρωνικού δικτύου, ένα δίκτυο τριών στρωμάτων επιλέχθηκε με συναρτήσεις ενεργοποίησης στους νευρώνες σιγμοειδούς ή λογιστικού τύπου. Αυτό το είδος νευρωνικού δικτύου δίνει συνήθως τα καλύτερα αποτελέσματα σε αυτούς τους τύπους ασκήσεων σύμφωνα με τη βιβλιογραφία που αναφέρεται στο παρόν άρθρο. Για να βρεθεί η τιμή ενοικίασης, λαμβάνεται ως είσοδος το ίδιο σύνολο μεταβλητών που καθορίζουν το ηδονικό μοντέλο και αυτό το σύνολο δεδομένων χρησιμοποιείται για την εκπαίδευση του μοντέλου.

Προηγουμένως, ορίστηκαν οι περιοχές εργασίας των νευρώνων, γεγονός που επηρεάζει θετικά τη λήψη ενός πιο αποδοτικού μοντέλου, καθώς δεν επιτρέπει στις περιοχές εργασίας να φύγουν από την περιοχή δεδομένων, διατηρώντας όλους τους νευρώνες σε λειτουργία.

Ως αποτέλεσμα όλων των παραπάνω, διαπιστώθηκε ότι η διαδικασία εκτίμησης μέσω νευρωνικού δικτύου είναι πολύ πιο ικανοποιητική από τη χρήση των δύο ηδονικών μοντέλων, με ακρίβεια 87% έναντι 46% που προέκυψε με το μοντέλο ηδονικής παλινδρόμησης. Επίσης, υπήρξε μείωση του μέσου τετραγωνικού σφάλματος, που έπεσε σε λιγότερο από 2.700 με το μοντέλο νευρωνικού δικτύου, επιβεβαιώνοντας ότι η εφαρμογή μοντέλων μηχανικής μάθησης επιτρέπει σε εταιρίες όπως η Airbnb να επιτύχουν καλύτερες διαδικασίες βελτιστοποίησης των τιμών από ό,τι με άλλα παραδοσιακά συστήματα.



Σχήμα 11. Σύγκριση των πραγματικών τιμών με τις προβλέψεις τιμών για 100 σπίτια.^[24]

5.2.4. Συμπεράσματα

Τα μοντέλα μηχανικής μάθησης σημειώνουν μεγάλη πρόοδο σε πολλά επιστημονικά πεδία, αλλά κάνουν επίσης τη διαφορά και σε επιχειρηματικό επίπεδο. Οι στρατηγικές που σχετίζονται με την επικοινωνία και το μάρκετινγκ, με τα νέα συστήματα gamma, την ερμηνεία των αντιδράσεων των χρηστών ή τα βελτιστοποιημένα μοντέλα τιμολόγησης, θεωρούνται σήμερα αδιανόητες από πολλές εταιρείες αν δεν συνδυαστούν με την τεχνητή νοημοσύνη και τα μοντέλα μεγάλων δεδομένων.

Στην περίπτωση της τουριστικής βιομηχανίας, η εμφάνιση της Airbnb τα τελευταία χρόνια έχει προσφέρει ένα ακόμη παράδειγμα για το πώς η μηχανική μάθηση μπορεί να βελτιστοποιήσει την απόδοση. Σε αντίθεση με τις αεροπορικές μεταφορές, όπου μόνο ένας πάροχος (η αεροπορική εταιρεία) ενεργεί με βάση τις πληροφορίες της ζήτησης, η Airbnb πρέπει να διαχειριστεί εκατομμύρια ιδιοκτήτες και πρακτικά τον ίδιο αριθμό τουριστών. Με άλλα λόγια, πρέπει να θέσει ως αποστολή της τη βελτιστοποίηση δύο βασικών στοιχείων: της τιμής ανάλογα με τις ιδιότητες του προϊόντος και της χρησιμότητας της ζήτησης για την παροχή της καλύτερης επιλογής στον χρήστη.

Οι ίδιοι οι υπάλληλοι της Airbnb αναγνωρίζουν σε διάφορα άρθρα που δημοσιεύονται στον ιστότοπό τους ότι χωρίς μοντέλα μηχανικής μάθησης θα ήταν αδύνατο να μεγιστοποιηθεί η απόδοση και αυτό το άρθρο, σε κάποιο βαθμό, δίνει προβολή σε αυτή τη δήλωση. Για τα

δεδομένα που επιλέχθηκαν (διαμερίσματα της Airbnb σε τρεις ισπανικούς τουριστικούς προορισμούς) δεν υπάρχει αμφιβολία ότι τα νευρωνικά δίκτυα λειτουργούν καλύτερα από τα παραδοσιακά μοντέλα παλινδρόμησης. Το αποτέλεσμα αυτό μας κάνει να σκεφτούμε πρώτα απ' όλα τις σχέσεις μεταξύ μεταβλητών που δεν είναι μόνο γραμμικές και, δεύτερον, τη μεγαλύτερη προβλεψιμότητα των μοντέλων μηχανικής μάθησης, η οποία αυξάνεται μόνο με τη συνεχή χρήση, τον όγκο των διαθέσιμων δεδομένων και τον αριθμό των μεταβλητών.

Εν ολίγοις, στο παράδειγμα που χρησιμοποιήθηκε για την ενοικίαση διαμερισμάτων στην Airbnb, τα νευρωνικά δίκτυα οδήγησαν σε σημαντική βελτίωση της πρόβλεψης σε σύγκριση με το ηδονικό μοντέλο τιμών. Σε επιχειρηματικό επίπεδο, αυτό μπορεί να σημαίνει ένα σημαντικό ανταγωνιστικό πλεονέκτημα, υπό την προϋπόθεση ότι τα οφέλη από την εφαρμογή του υπερκαλύπτουν το κόστος ανάπτυξής του.

5.3. Γενικά συμπεράσματα και μελλοντικές κατευθύνσεις

Στην πρώτη έρευνα του συγκεκριμένου κεφαλαίου, οι ερευνητές περιέγραψαν τις προκλήσεις του προβλήματος της δυναμικής τιμολόγησης στην Airbnb και παρουσίασαν ένα μοντέλο στρατηγικής τιμολόγησης το οποίο αναπτύχθηκε με σκοπό να βοηθήσει τους οικοδεσπότες της Airbnb ώστε να καθορίζουν τις τιμές των καταλυμάτων τους αποτελεσματικότερα. Τα αποτελέσματα αξιολόγησης εκτός σύνδεσης και σε σύνδεση δείχνουν ότι το προτεινόμενο μοντέλο στρατηγικής GBM αποδίδει σημαντικά καλύτερα από ένα άμεσο στρατηγικό μοντέλο τιμολόγησης max-rev.

Στη δεύτερη έρευνα η οποία πραγματοποιήθηκε για την περίπτωση της Airbnb στην Κοινότητα της Βαλένθια οι ερευνητές συμπεραίνουν πως για τα δεδομένα που επιλέχθηκαν δεν υπάρχει καμία αμφιβολία ότι τα νευρωνικά δίκτυα λειτουργούν καλύτερα από τα παραδοσιακά μοντέλα παλινδρόμησης. Οδήγησαν σε σημαντική βελτίωση της πρόβλεψης σε σύγκριση με το ηδονικό μοντέλο τιμών.

Σε μελλοντικές έρευνες, οι ερευνητές θα μπορούσαν να εργαστούν για να βελτιώσουν την καμπύλη ζήτησης, ώστε να υπάρχει η δυνατότητα να επανεξεταστεί η άμεση στρατηγική μεγιστοποίησης εσόδων στο μέλλον. Επίσης, αφού βλέπουμε ότι η χρήση των νευρωνικών δικτύων βελτιστοποιεί τα αποτελέσματα, οι μελλοντικοί ερευνητές θα μπορούσαν να επικεντρωθούν στη χρησιμοποίηση και τη βελτίωση του συγκεκριμένου μοντέλου.

6. ΙΚΑΝΟΠΟΙΗΣΗ ΤΩΝ ΕΠΙΣΚΕΠΤΩΝ

6.1. Ένα πελατοκεντρικό σύστημα αξιολόγησης για την αξιολόγηση της ικανοποίησης των επισκεπτών της Airbnb με χρήση Μηχανικής Μάθησης και NLP.^[25]

6.1.1. Εισαγωγή

Η ακόλουθη έρευνα πραγματοποιήθηκε από τους Mohamed Chiny, Omar Bencharef, Moulay Youssef Hadi και Younes Chihab στο Μαρόκο και δημοσιεύθηκε το Φεβρουάριο του 2021.

Στην τουριστική βιομηχανία, η ευχαρίστηση του πελάτη είναι το σημαντικότερο κομμάτι. Αν υπάρχει ευχαρίστηση από τη μεριά του πελάτη, τότε χτίζονται ισχυροί δεσμοί εμπιστοσύνης και βελτιώνονται κατά πολύ τα οικονομικά οφέλη. Η παρούσα εργασία επιχειρεί να κατανοήσει τη σχέση μεταξύ των στοιχειωδών βαθμολογιών της Airbnb και της συνολικής βαθμολογίας ανά κατηγορία πελατών, εξετάζοντας δεδομένα σχετικά με ακίνητα που βρίσκονται στο Λονδίνο και τα οποία έχουν συλλεχθεί από την ιστοσελίδα Inside Airbnb. Οι βαθμολογίες και οι απόψεις που συλλέχθηκαν ολοκληρώθηκαν μεταξύ Δεκεμβρίου 2009 και Απριλίου 2020.

Αφού τα υπάρχοντα δεδομένα δεν ταξινομούνται ανά κατηγορία πελατών, πρέπει να εξεταστούν οι απόψεις που αφήνουν οι ίδιοι οι πελάτες προχωρώντας σε εξόρυξη κειμένου. Πράγματι, μετά τον καθαρισμό και το φιλτράρισμα των συλλεχθέντων δεδομένων, από τα οποία προέκυψαν 100.000 δεδομένα έτοιμα προς ανάλυση, πραγματοποιείται στη συνέχεια η τμηματοποίησή τους προκειμένου να εξαχθεί η κατηγορία των πελατών από τις απόψεις τους με τη χρήση αλγορίθμων επεξεργασίας φυσικής γλώσσας (NLP). Λόγω του μεγάλου όγκου των προς επεξεργασία δεδομένων καθώς και του σημασιολογικού τους πλούτου, υιοθετείται μια προσέγγιση βασισμένη στη μηχανική μάθηση. Η ανάλυση των δεδομένων πραγματοποιείται στη συνέχεια με τη χρήση της πολλαπλής γραμμικής παλινδρόμησης και της διανυσματικής μηχανής υποστήριξης για αλγόριθμους παλινδρόμησης.

Με σκοπό να επικυρωθούν τα δύο μοντέλα, οι ερευνητές εκπαίδευσαν ένα τεχνητό νευρωνικό δίκτυο για την πρόβλεψη της συνολικής βαθμολογίας από τις συνολικές βαθμολογίες ενός δείγματος δεδομένων το οποίο αποτελεί το αντικείμενο της μελέτης.

Τα αποτελέσματα αυτής της μελέτης θα είναι σημαντικά για να βοηθήσουν τις πλατφόρμες φιλοξενίας P2P (στην προκειμένη περίπτωση, την Airbnb) καθώς και τους οικοδεσπότες που προσφέρουν τα προϊόντα τους εκεί, να κατανοήσουν καλύτερα τις ανάγκες των πελατών τους, ώστε να αναπτύξουν την εμπειρία τους κατά τη διάρκεια της διαμονής τους. Αντίθετα, θα είναι χρήσιμη για την Airbnb ώστε να δημιουργήσει ένα εξατομικευμένο σύστημα βαθμολόγησης που θα λαμβάνει υπόψη την κατηγορία των πελατών. Με άλλα λόγια, για τους μελλοντικούς πελάτες που αναζητούν κατάλυμα στην πλατφόρμα της Airbnb, θα προβάλλονται

συγκεκριμένα οι βαθμολογίες και οι κριτικές που αφορούν την κατηγορία τους. Αυτό θα τους βοηθήσει ώστε να πάρουν μια απόφαση σε σχέση με τις προσφορές που προβάλλονται.

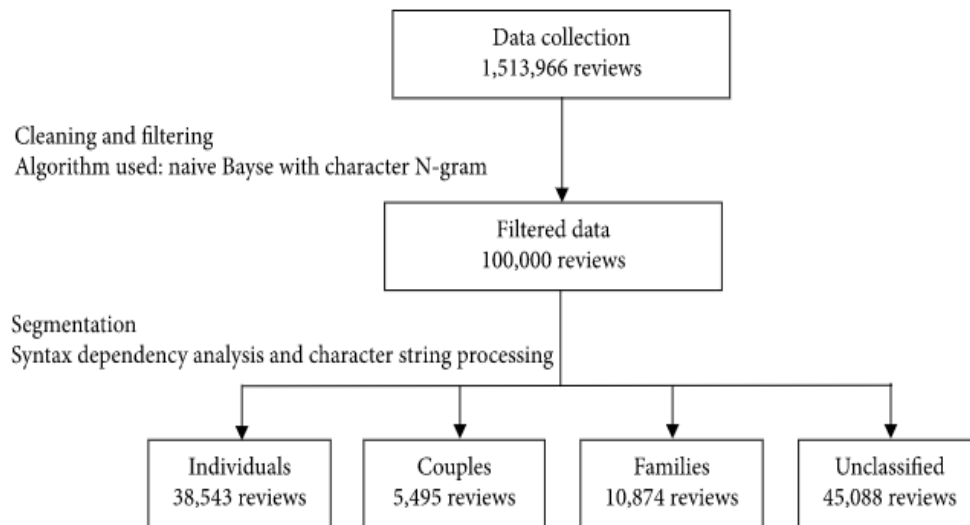
6.1.2. Χαρακτηριστικά και Μέθοδοι

Οι κριτικές και οι αξιολογήσεις των πελατών της Airbnb συλλέγονται ως μέρος της διαδικασίας της συγκεκριμένης έρευνας. Αυτά τα αρχεία στη συνέχεια καθαρίζονται και φιλτράρονται ώστε να διατηρούνται μόνο οι κριτικές οι οποίες είναι γραμμένες στα αγγλικά. Στη συνέχεια, τμηματοποιούνται προκειμένου να βρεθεί η ομάδα που ανήκει ο πελάτης. Τέλος, εκπαιδεύτηκαν δύο μοντέλα παλινδρόμησης για τον προσδιορισμό των βαθμολογιών με βάση τους δείκτες ικανοποίησης των πελατών πριν χρησιμοποιηθεί ένα νευρωνικό δίκτυο για την πρόβλεψη των βαθμολογιών παράλληλα με τις δύο τεχνικές που χρησιμοποιήθηκαν προηγουμένως.

Τα δεδομένα για την παρούσα μελέτη συλλέχθηκαν από την Inside Airbnb, η οποία είναι ένας ιστότοπος έρευνας που περιέχει δεδομένα σχετικά με την ενοικίαση ακινήτων στην Airbnb. Τα δεδομένα αυτά, τα οποία μεταφέρουν αρκετές πληροφορίες, αφορούν κατοικίες που βρίσκονται στο Λονδίνο. Εκτός από τα καταλύματα και τα λεπτομερή χαρακτηριστικά τους (όπως ο τύπος, η περιοχή και το όνομα του οικοδεσπότη), τα δεδομένα περιέχουν επίσης το ημερολόγιο των κρατήσεων και τις κριτικές των πελατών που έκαναν κράτηση σε αυτά τα ακίνητα μεταξύ Δεκεμβρίου 2009 και Απριλίου 2020. Το σύνολο των δεδομένων που συλλέχθηκαν περιλαμβάνει 86.357 μονάδες κατοικίας που κατανέμονται σε 32 συνοικίες και 1.513.966 κριτικές που άφησαν οι πελάτες.

Για να εξασφαλιστεί ένα αποδεκτό επίπεδο συνάφειας των δεδομένων που συλλέγονται, διατηρούνται μόνο οι γνώμες των πελατών που δεν παρουσιάζουν λειψά ή ασυνεπή πεδία. Στη συνέχεια, φιλτράρονται οι κριτικές που γράφτηκαν στα αγγλικά. Αυτό θα μπορούσε να γίνει χρησιμοποιώντας αναγνώριση γλώσσας με βάση την αναγνώριση λέξεων από λεξικό, ωστόσο, η μέθοδος αυτή απαιτεί ένα μεγάλο λεξικό και επιπλέον σημαντικό χρόνο επεξεργασίας. Ως εκ τούτου, χρησιμοποιείται η προσέγγιση του υπολογισμού της πιθανότητας της γλώσσας με βάση τα χαρακτηριστικά της προφοράς χρησιμοποιώντας τον αλγόριθμο Naive Bayes με χαρακτήρα N-gram. Η μέθοδος αυτή έχει αποδείξει την απόδοσή της με ακρίβεια 99,8% και έχει ικανοποιητική ταχύτητα ανίχνευσης.

Παρά το γεγονός ότι η μέθοδος που χρησιμοποιείται για τον εντοπισμό της γλώσσας είναι αρκετά γρήγορη, αλλά δεδομένου του αριθμού των προς ανάλυση γνώμων (1.513.966 γνώμες), ο χρόνος υπολογισμού που απαιτείται για την εκτέλεση αυτής της επεξεργασίας εξακολουθεί να είναι υψηλός. Ως εκ τούτου, διατηρήθηκαν 100.000 κριτικές έτοιμες να χρησιμοποιηθούν, οι οποίες φαίνονται και στο Σχήμα 12.



Σχήμα 12. Διαδικασία καθαρισμού, φιλτραρίσματος και τμηματοποίησης των δεδομένων που συλλέγονται από την ιστοσελίδα Inside Airbnb.^[25]

6.1.3. Αποτελέσματα

Τα συνολικά δεδομένα χωρίστηκαν σε δύο παρτίδες, δεδομένα εκπαίδευσης και δεδομένα δοκιμής, με ποσοστά 80% και 20% αντίστοιχα. Η επιλογή των δεδομένων για την εκπαίδευση και τη δοκιμή ήταν εντελώς τυχαία.

Μετά την εκπαίδευση των δύο μοντέλων, οι ερευνητές μπόρεσαν να λάβουν τα ακριβή αποτελέσματα του Πίνακα 9 χρησιμοποιώντας το R2. Οι δύο αλγόριθμοι που χρησιμοποιήθηκαν απεικονίζουν αρκετά κοντινές ακρίβειες για κάθε μία από τις τρεις κατηγορίες που μελετήθηκαν.

	MLR	SVR
Individual	0.660	0.657
Couple	0.684	0.676
Family	0.712	0.711

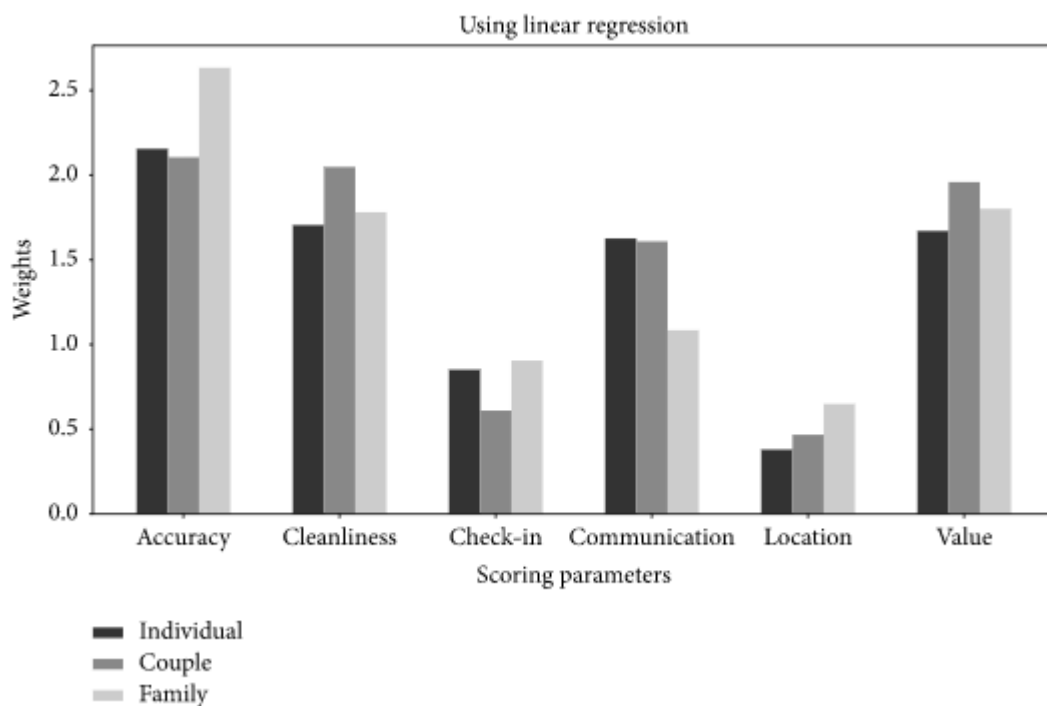
Πίνακας 9. Ακρίβεια των αλγορίθμων MLR και SVR.^[25]

Ο Πίνακας 10 απεικονίζει τους συντελεστές που υπολογίστηκαν με τη χρήση των αλγορίθμων MLR και SVR σύμφωνα με τις προκαθορισμένες κατηγορίες πελατών. Υπολογίστηκε επίσης ο μέσος όρος αυτών των συντελεστών, διότι υπολογίζεται ότι θα μας φανούν χρήσιμοι κατά τις ακόλουθες ερμηνείες.

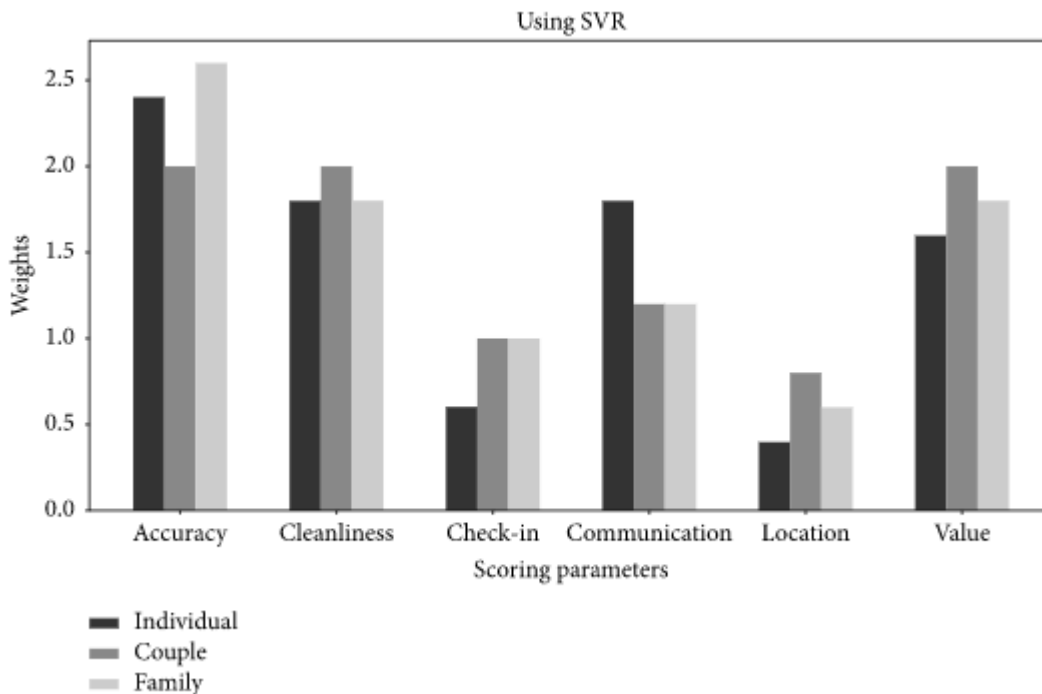
	MLR algorithm				SVR algorithm			
	Individual	Couple	Family	Mean	Individual	Couple	Family	Mean
Accuracy	2.155	2.106	2.634	2.298	2.402	2.000	2.600	2.334
Cleanliness	1.705	2.048	1.781	1.845	1.799	2.000	1.800	1.866
Check-in	0.852	0.610	0.905	0.789	0.602	1.000	1.000	0.867
Communication	1.626	1.609	1.084	1.440	1.799	1.200	1.199	1.399
Location	0.380	0.467	0.649	0.499	0.398	0.800	0.600	0.599
Value	1.670	1.959	1.802	1.810	1.598	2.000	1.800	1.799

Πίνακας 10. Συντελεστές που ενεργούν στους 6 δείκτες ικανοποίησης με τη χρήση των αλγορίθμων MLR και SVR.^[25]

Τα σχήματα 13 και 14 αντίστοιχα, απεικονίζουν την κατανομή των συντελεστών με τη χρήση των αλγορίθμων MLR και SVR.



Σχήμα 13. Κατανομή των συντελεστών που υπολογίζονται με τον αλγόριθμο MLR.^[25]



Σχήμα 14. Κατανομή των συντελεστών που υπολογίζονται με τον αλγόριθμο SVR.^[25]

6.1.4. Συμπεράσματα

Πολλές μελέτες έχουν ενδιαφερθεί να απομυθοποιήσουν τις διαστάσεις της ικανοποίησης των πελατών όσον αφορά τον τομέα της φιλοξενίας P2P κατά την τελευταία δεκαετία, δεδομένου ότι η Airbnb είναι σήμερα η μεγαλύτερη πλατφόρμα για καταλύματα βραχυχρόνιας μίσθωσης και έχει διαταράξει σημαντικά το παραδοσιακό σύστημα κρατήσεων.

Στη συγκεκριμένη έρευνα, επιδιώχθηκε να κατανοηθεί ποιοι δείκτες καθορίζουν την ικανοποίηση των πελατών, λαμβάνοντας υπόψη την κατηγορία τους σε σχέση με τα καταλύματα που προσφέρει η Airbnb. Συγκεντρώθηκε, λοιπόν, ένα μεγάλο πλήθος κριτικών από χρήστες που έκαναν κράτηση ακινήτων στο Λονδίνο μεταξύ Δεκεμβρίου 2009 και Απριλίου 2020 από την ιστοσελίδα Inside Airbnb. Μετά τον καθαρισμό, το φιλτράρισμα και την τμηματοποίηση αυτών των απόψεων, χρησιμοποιώντας κυρίως αλγόριθμους επεξεργασίας φυσικής γλώσσας (NLP), χρησιμοποιήθηκαν αλγόριθμοι παλινδρόμησης, στην προκειμένη περίπτωση, πολλαπλή γραμμική παλινδρόμηση (MLR) και παλινδρόμηση διανυσμάτων υποστήριξης (SVR), προκειμένου να υπολογιστούν οι συντελεστές που δρουν στις στοιχειώδεις βαθμολογίες που σημείωσαν οι πελάτες και οι οποίοι επηρεάζουν τη συνολική βαθμολογία.

Τα αποτελέσματα υποδηλώνουν ότι οι πελάτες κατατάσσουν τους συγκεκριμένους δείκτες με την ακόλουθη σειρά: ακρίβεια, καθαριότητα, αξία, επικοινωνία, check-in και τοποθεσία.

Επιπλέον, η κατανόηση αυτών των δεικτών αλλάζει από τη μία κατηγορία στην άλλη, για παράδειγμα, οι οικογένειες αποδίδουν περισσότερη σημασία στην ακρίβεια περισσότερο από άλλες κατηγορίες, τα ζευγάρια από την άλλη, ενδιαφέρονται, εκτός από την ακρίβεια, για την καθαριότητα και την αξία και η επικοινωνία είναι ένας δείκτης που λαμβάνεται σοβαρά υπόψη από όλα τα άτομα.

6.2. Γενικά συμπεράσματα και μελλοντικές κατευθύνσεις

Συμπερασματικά, οι συγγραφείς της συγκεκριμένης έρευνας πιστεύουν ότι η έρευνα αυτή μπορεί να συμβάλει στην καλύτερη κατανόηση των προσδοκιών των πελατών στον τομέα της ομότιμης φιλοξενίας, ιδίως μέσω της Airbnb, προκειμένου να γίνουν κατανοητοί λαμβάνοντας υπόψη την κατηγορία στην οποία ανήκουν, ώστε να τους παραχθεί η καλύτερη δυνατή εμπειρία. Επιπρόσθετα, η Airbnb μπορεί να εξετάσει το ενδεχόμενο δημιουργίας ενός νέου συστήματος βαθμολόγησης που θα λαμβάνει υπόψη την κατηγορία των πελατών, το οποίο πιθανότατα θα βοηθήσει τους μελλοντικούς καταναλωτές να κάνουν καλύτερες επιλογές όσον αφορά τις προσφορές που παρατίθενται. Αντίθετα, πρόσφατες έρευνες συμβάλλουν στην κατανόηση της βιβλιογραφίας σχετικά με τους παράγοντες που επηρεάζουν την ικανοποίηση των πελατών, ιδίως στο πλαίσιο της ομότιμης φιλοξενίας (P2P).

7. ΕΠΙΛΟΓΟΣ

Η μηχανική μάθηση, είναι ένα υποπεδίο της επιστήμης των υπολογιστών και πιο συγκεκριμένα της τεχνητής νοημοσύνης. Αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη και διερευνά τη μελέτη και την κατασκευή αλγορίθμων οι οποίοι μπορούν να μαθαίνουν από τα δεδομένα τους και να κάνουν προβλέψεις σχετικά με αυτά.

Η Airbnb είναι ιστοσελίδα καταχώρισης, εύρεσης και ενοικίασης καταλυμάτων. Ιδρύθηκε τον Αύγουστο του 2008 από τον Joe Gebbia, τον Brian Chesky και τον Nathan Blecharczyk και η βάση της βρίσκεται στο Σαν Φρανσίσκο της Καλιφόρνια. Μέσω της Airbnb, ο επισκέπτης μπορεί να βρει μια μεγάλη ποικιλία καταλυμάτων όπως διαμέρισμα, κάστρο, βίλα κ.α., ανάλογη με το μέγεθος της κοινότητας των χρηστών της, οι οποίοι υπόσχονται μοναδικές ταξιδιωτικές εμπειρίες.

Στη συγκεκριμένη βιβλιογραφική εργασία, έγινε εξόρυξη γνώσης από κείμενα τουριστικών εφαρμογών και εκτεταμένη βιβλιογραφική ανάλυση των αλγορίθμων μηχανικής μάθησης που χρησιμοποιούνται σε διάφορους τομείς της πλατφόρμας της Airbnb, όπως η πρόβλεψη τιμών, η αναζήτηση, η δυναμική τιμολόγηση και η ικανοποίηση των επισκεπτών. Αναλύθηκαν επίσης διάφορα χαρακτηριστικά που βοήθησαν στην εκπαίδευση των αλγορίθμων.

Μερικά από τα μοντέλα που χρησιμοποιήθηκαν για την ανάλυση των εργασιών είναι η γραμμική παλινδρόμηση, τα νευρωνικά δίκτυα, tree-based μοντέλα, SVR, GBDT μοντέλα όπως το XGBoost κ.α., με έμφαση κυρίως στο XGBoost και τα νευρωνικά δίκτυα, αφού στις περισσότερες έρευνες που αναλύθηκαν δίνουν τα καλύτερα αποτελέσματα.

8. ΒΙΒΛΙΟΓΡΑΦΙΑ

1. ΒΙΚΙΠΑΙΔΕΙΑ. (2019). *Airbnb*. <https://el.wikipedia.org/wiki/Airbnb>
2. Airbnb. (2015). *Shared Opportunity: How Airbnb Benefits Communities*. https://www.ftc.gov/system/files/documents/public_comments/2015/05/01740-96152.pdf
3. David Curry. (May 6, 2021). *Airbnb Revenue and Usage Statistics*. <https://www.businessofapps.com/data/airbnb-statistics/>
4. Brown, M., 2016. *The Making of Airbnb*. *Boston Hospitality Review*, 4(1), pp. 33-38. https://www.bu.edu/bhr/files/2016/02/The-Making-of-Airbnb_Brown_Winter2016.pdf
5. Airbnb. (2016). *Εγγραφείτε για δωρεάν επαγγελματική φωτογράφιση*. <https://www.airbnb.gr/d/pro-photography>
6. Airbnb. (2016). *Φιλοξενήστε στο χώρο σας και βγάλτε χρήματα με την Airbnb*. <https://www.airbnb.gr/host/homes>
7. Ινστιτούτο του Συνδέσμου Ελληνικών Τουριστικών Επιχειρήσεων (ΙΝΣΕΤΕ). (2018-2019). *Η ακτινογραφία της Airbnb στην Ελλάδα - Πάνω από 1 δισ. ευρώ τα ετήσια έσοδα από τις μισθώσεις σπιτιών*. <https://www.tornosnews.gr/foreis/sete/40914-insete-pano-apo-1-dis-eyro-ta-ethsia-esoda-apo-tis-misthoseis-spition-sthn-ellada.html>
8. Geetha R. και Bhanu Sree Reddy D. (2018) *Recruitment Through Artificial Intelligence: A Conceptual Study*. *International Journal of Mechanical Engineering and Technology*. Τόμ. 9, 7. https://www.researchgate.net/publication/326929416_Recruitment_through_artificial_intelligence_A_conceptual_study
9. De Spiegeleire Stephan, Maas Matthijs και Sweijts Tim. (2017) *Artificial Intelligence and the Future of Defense*. s.l. : HCSS, 2017. 978-94-92102-54-6.
10. ΒΙΚΙΠΑΙΔΕΙΑ. (2021). *Μηχανική Μάθηση*. https://el.wikipedia.org/wiki/%CE%9C%CE%B7%CF%87%CE%B1%CE%BD%CE%B9%CE%BA%CE%AE_%CE%BC%CE%AC%CE%B8%CE%B7%CF%83%CE%B7
11. Mitchell, T. (1997). *Machine Learning*, McGraw Hill, *Machine Learning*, McGraw Hill, p.2
12. Harnad, Stevan (2008), «*The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence*», στο: Epstein, Robert; Peters, Grace, επιμ., *The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Kluwer. <https://eprints.soton.ac.uk/262954/>
13. Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning*.

14. Kalehbasti, P. R., Nikolenko, L., & Rezaei, H. (2019). *Airbnb Price Prediction Using Machine Learning and Sentiment Analysis*. https://doi.org/10.1007/978-3-030-84060-0_11
15. Dhillon, J., Eluri, N. P., Kaur, D., Chhipa, A., Gadupudi, A., Eravi, R. C., & Pirouz, M. (2021). Analysis of Airbnb Prices using Machine Learning Techniques. *2021 IEEE 11th Annual Computing and Communication Workshop and Conference, CCWC 2021*, 297–303. <https://doi.org/10.1109/CCWC51732.2021.9376144>
16. Tang, E., & Sangani, K. (n.d.). *Neighborhood and Price Prediction for San Francisco Airbnb Listings*. http://cs229.stanford.edu/proj2015/236_report.pdf
17. Luo, Y., Zhou, X., & Zhou, Y. (2019). *Predicting Airbnb Listing Price Across Different Cities*. http://cs229.stanford.edu/proj2019aut/data/assignment_308832_raw/26647491.pdf
18. Yang, S. (2021). Learning-based Airbnb Price Prediction Model. *Proceedings - 2nd International Conference on E-Commerce and Internet Technology, ECIT 2021*, 283–288. <https://doi.org/10.1109/ECIT52743.2021.00068>
19. Haldar, M., Abdool, M., Ramanathan, P., Xu, T., Yang, S., Duan, H., Zhang, Q., Barrow-Williams, N., Turnbull, B. C., Collins, B. M., & Legrand, T. (2019). Applying deep learning to airbnb search. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1927–1935. <https://doi.org/10.1145/3292500.3330658>
20. Grbovic, M. (2017). Search ranking and personalization at Airbnb. *RecSys 2017 - Proceedings of the 11th ACM Conference on Recommender Systems*, 339–340. <https://doi.org/10.1145/3109859.3109920>
21. Turnbull, B. C. (2019). Learning intent to book metrics for Airbnb search. *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, 3265–3271. <https://doi.org/10.1145/3308558.3313648>
22. Grbovic, M., & Cheng, H. (2018). Real-time personalization using embeddings for search ranking at Airbnb. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 311–320. <https://doi.org/10.1145/3219819.3219885>
23. Ye, P., Wu, C. H., Qian, J., Zhou, Y., Chen, J., de Mars, S., Yang, F., & Zhang, L. (2018). Customized regression model for Airbnb dynamic pricing. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 932–940. <https://doi.org/10.1145/3219819.3219830>
24. Moreno-Izquierdo, L., Egorova, G., Peretó-Rovira, A., & Más-Ferrando, A. (2018). Exploring the use of artificial intelligence in price maximisation in the tourism sector: its application in the case of Airbnb in the Valencian Community. In *Investigaciones*

Regionales-Journal of Regional Research (Vol. 42).

<https://www.researchgate.net/publication/332320494> Exploring the use of artificial intelligence in price maximisation in the tourism sector Its application in the case of Airbnb in the Valencian Community

25. Chiny, M., Bencharef, O., Hadi, M. Y., & Chihab, Y. (2021). A Client-Centric Evaluation System to Evaluate Guest's Satisfaction on Airbnb Using Machine Learning and NLP. *Applied Computational Intelligence and Soft Computing*, 2021.
<https://doi.org/10.1155/2021/6675790>
26. Zervas, G., Proserpio, D., & Byers, J. W. (2021). A first look at online reputation on Airbnb, where every stay is above average. *Marketing Letters*, 32(1).
<https://doi.org/10.1007/s11002-020-09546-4>
27. Corporate Finance Institute. *Random Forest. A combination of decision trees that can be modeled for prediction and behavior analysis*.
<https://corporatefinanceinstitute.com/resources/knowledge/other/random-forest/>