

Εισαγωγή στα Big Data & Analytics

Εισαγωγή στην Επιστήμη Δεδομένων

Εισαγωγή στην Επιστήμη Δεδομένων

Οι επιχειρήσεις, οι τράπεζες, οι κυβερνήσεις, οι ιστότοποι κοινωνικής δικτύωσης, οι οργανισμοί αναφοράς πιστώσεων και οι εταιρείες τηλεπικοινωνιών, μεταξύ άλλων, **κατέχουν τεράστιες ποσότητες πληροφοριών που μας αφορούν.**

Ξέρουν

- πού ζούμε,
- τι ξοδεύουμε
- τα χρήματά μας,
- ποιοι είναι οι φίλοι και η οικογένειά μας,
- οι συμπάθειες και οι αντιπάθειες μας,
- ο τρόπος ζωής μας και οι απόψεις μας.

•

Εισαγωγή στην Επιστήμη Δεδομένων

Κάθε χρόνο η ποσότητα των ηλεκτρονικών πληροφοριών για εμάς αυξάνεται καθώς χρησιμοποιούμε όλο και περισσότερο τις

- υπηρεσίες διαδικτύου,
- τα κοινωνικά μέσα και
- τις έξυπνες συσκευές

για να μεταφέρουμε όλο και περισσότερο τη ζωή μας στο διαδικτυακό περιβάλλον.

Εισαγωγή στην Επιστήμη Δεδομένων

Μέχρι τις αρχές της δεκαετίας του 2000, η πρωταρχική πηγή των ατομικών (καταναλωτικών) δεδομένων ήταν τα ηλεκτρονικά αποτυπώματα που αφήσαμε πίσω καθώς κινούσαμε μέσα από τη ζωή, όπως

- **Οι συναλλαγές με πιστωτικές κάρτες,**
- **Οι ηλεκτρονικές αγορές και**
- **Τα αιτήματα για ασφαλιστικές προσφορές.**

Αυτές οι πληροφορίες απαιτούνται για τη δημιουργία λογαριασμών, τη διατήρηση λογαριασμών μέχρι κάποια συγκεκριμένη ημερομηνία, και να παρέχει έναν έλεγχο των συναλλαγών που έχουν συμβεί μεταξύ των παρόχων υπηρεσιών και των πελατών τους.

Εισαγωγή στην Επιστήμη Δεδομένων

Τα τελευταία χρόνια οι οργανισμοί ενδιαφέρονται όλο και περισσότερο για τους χώρους μεταξύ των συναλλαγών μας και των μονοπατιών που μας οδήγησαν στις αποφάσεις που πήραμε. Καθώς κάνουμε περισσότερα πράγματα ηλεκτρονικά, έχουν γίνει διαθέσιμες οι πληροφορίες που δίνουν γνώσεις σχετικά με τις διαδικασίες σκέψης μας και τις επιρροές που μας οδήγησαν να συμμετάσχουμε σε μια δραστηριότητα και όχι σε μια άλλη.

Εισαγωγή στην Επιστήμη Δεδομένων

Ένας έμπορος μπορεί να κατανοήσει γιατί αγοράσαμε ένα προϊόν από αυτόν και όχι από έναν ανταγωνιστή, εξετάζοντας ποια διαδρομή κάναμε πριν το αγοράσαμε-

- Ποιες ιστοσελίδες επισκεφθήκαμε;
- Ποια άλλα προϊόντα εξετάσαμε;
- Ποιες κριτικές συμβουλευτήκαμε;

Ομοίως, τα κοινωνικά μέσα παρέχουν κάθε είδους πληροφορίες για τον πελάτη (τι σκεφτόμαστε, με ποιον μιλάμε και για τι μιλάμε) και τα τηλέφωνα μας. Άλλες συσκευές παρέχουν πληροφορίες για το πού βρισκόμαστε.

Εισαγωγή στην Επιστήμη Δεδομένων

Όλες αυτές οι πληροφορίες για τους ανθρώπους είναι εξαιρετικά χρήσιμες για πολλούς και διαφορετικούς λόγους, αλλά μια ενδιαφέρουσα εφαρμογή ειδικότερα είναι η **πρόβλεψη μελλοντικής συμπεριφοράς**.

Χρησιμοποιώντας πληροφορίες σχετικά με τον τρόπο ζωής, τις κινήσεις και τις προηγούμενες συμπεριφορές των ανθρώπων, οι εταιρείες μπορούν να προβλέψουν για τους πελάτες τους

- τι είναι πιθανό να κάνουν,
- πότε θα το κάνουν και
- πού θα συμβεί αυτή η δραστηριότητα.

Εισαγωγή στην Επιστήμη Δεδομένων

Στη συνέχεια χρησιμοποιούν αυτές τις προβλέψεις για να προσαρμόσουν τον τρόπο με τον οποίο αλληλεπιδρούν με τους πελάτες.

Ο λόγος που γίνεται αυτό είναι για να επηρεάσουν τη συμπεριφορά των ανθρώπων, προκειμένου να μεγιστοποιήσουν την αξία των σχέσεων που έχουν μαζί τους.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση;

Μέχρι τη δεκαετία του 1980, πολλοί οργανισμοί βρέθηκαν με βάσεις δεδομένων πελατών που είχαν αυξηθεί στο σημείο όπου ο όγκος των δεδομένων που κατείχαν είχε γίνει πολύ μεγάλος για να μπορούν να το αναλύσουν με παραδοσιακές τεχνικές.

Ο όρος "**εξόρυξη δεδομένων**" επινοήθηκε για να περιγράψει μια σειρά αυτοματοποιημένων τεχνικών που θα μπορούσαν να εφαρμοστούν για να διερευνήσουν αυτές τις βάσεις δεδομένων και να βγάλουν χρήσιμα συμπεράσματα για το τι σημαίνουν τα δεδομένα.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση;

Εάν θέλετε έναν συνοπτικό ορισμό της εξόρυξης δεδομένων, τότε "**η ανάλυση μεγάλων και σύνθετων συνόλων δεδομένων**" είναι ένα καλό μέρος για να ξεκινήσετε.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση;

Πολλά από τα εργαλεία που χρησιμοποιούνται για την εξόρυξη δεδομένων είναι τυποποιημένες **στατιστικές μέθοδοι** που υπάρχουν εδώ και δεκαετίες, όπως η **γραμμική παλινδρόμηση** και η **ομαδοποίηση**.

Ωστόσο, η εξόρυξη δεδομένων περιλαμβάνει επίσης ένα ευρύ φάσμα άλλων τεχνικών για την ανάλυση δεδομένων που προέκυψαν από την έρευνα για την **τεχνητή νοημοσύνη** (μηχανική μάθηση), την **εξελικτική πληροφορική** τη **θεωρία των παιγνίων**, κ.λπ.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση;

Η εξόρυξη δεδομένων είναι ένα πολύ ευρύ θέμα, που χρησιμοποιείται για πολλά πράγματα.

- Η ανίχνευση μοτίβων στα δορυφορικά δεδομένα,
- η πρόβλεψη των κινήσεων των τιμών των μετοχών,
- η αναγνώριση προσώπου και
- η πρόβλεψη της κυκλοφοριακής συμφόρησης

είναι μερικά μόνο παραδείγματα όπου εφαρμόζεται συνήθως η εξόρυξη δεδομένων.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση;

Ωστόσο, η πιο παραγωγική χρήση της εξόρυξης δεδομένων είναι να προσδιοριστούν οι σχέσεις σε δεδομένα που δίνουν μια εικόνα για τις ατομικές προτιμήσεις και, το σημαντικότερο, τι είναι πιθανό να κάνει κάποιος σε ένα δεδομένο σενάριο.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση;

Αυτό είναι σημαντικό επειδή εάν μια επιχείρηση γνωρίζει τι είναι πιθανό να κάνει κάποιος ενδεχόμενος πελάτης, τότε μπορεί να προσαρμόσει την πολιτική της του για να μεγιστοποιήσει τους δικούς της στόχους.

Για τις επιχειρήσεις ο στόχος είναι συνήθως η μεγιστοποίηση των κερδών.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση;

Ωστόσο, και η κυβερνήσεις καθώς και άλλες μη κερδοσκοπικές οργανώσεις έχουν επίσης λόγους να θέλουν να μάθουν πώς θα συμπεριφέρονται οι άνθρωποι και στη συνέχεια να αναλάβουν δράση για να την αλλάξουν ή να την αποτρέψουν.

Για παράδειγμα, **οι φορολογικές αρχές** θέλουν να προβλέψουν ποιος είναι απίθανο να καταθέσει σωστά τη φορολογική του δήλωση και, ως εκ τούτου, να στοχεύσει αυτά τα άτομα για δράση από φορολογικούς επιθεωρητές.

Ομοίως, **τα πολιτικά κόμματα** θέλουν να εντοπίσουν τους μη σταθερούς ψηφοφόρους και στη συνέχεια να τους ωθήσουν, χρησιμοποιώντας ατομικά προσαρμοσμένες επικοινωνίες, να τους ψηφίσουν.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση;

Κάποια στιγμή στα μέσα της δεκαετίας του 2000 ο όρος "**προγνωστική ανάλυση**" (**Predictive analytics**) έγινε συνώνυμος με τη χρήση της εξόρυξης δεδομένων για την ανάπτυξη εργαλείων για την πρόβλεψη της συμπεριφοράς των ατόμων.

Επομένως, τα προγνωστικά αναλυτικά στοιχεία είναι απλώς ένας όρος που χρησιμοποιείται για να περιγράψει την εφαρμογή της εξόρυξης δεδομένων σε αυτό το είδος προβλήματος.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση; (παράδειγμα)

Για την κατασκευή ενός μοντέλου βαθμολόγησης πιστοληπτικής ικανότητας, η προγνωστική ανάλυση χρησιμοποιείται για την ανάλυση δεδομένων από χιλιάδες ιστορικές δανειακές συμβάσεις για να προσδιορίσει ποια χαρακτηριστικά των δανειοληπτών ήταν ενδεικτικά ότι είναι "καλοί" πελάτες που αποπλήρωσαν τα δάνειά τους ή "κακοί" πελάτες που αθέτησαν.

Έχοντας χρησιμοποιήσει προγνωστικά αναλυτικά στοιχεία για την κατασκευή ενός μοντέλου, μπορεί κανείς να χρησιμοποιήσει το μοντέλο για να κάνει προβλέψεις σχετικά με τη μελλοντική συμπεριφορά αποπληρωμής των νέων αιτούντων δάνειο

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση; (παράδειγμα)

Στην Αμερική πολλά πιστωτικά ιδρύματα χρησιμοποιούν ένα μοντέλο πιστοληπτικής ικανότητας που αναπτύχθηκε από την **FICO Corporation** (πρώην Fair Isaac Corporation)

Συνήθως, οι βαθμολογίες FICO κυμαίνονται από περίπου 300 έως περίπου 850.

Όσο υψηλότερη είναι η βαθμολογία σας τόσο πιο αξιόπιστη είστε.

Παρόμοιες βαθμολογίες χρησιμοποιούνται από οργανισμούς σε όλο τον κόσμο.

Ένα παράδειγμα ενός μοντέλου βαθμολόγησης πίστωσης είναι το ακόλουθο

Constant	670		
Employment status		Outstanding mortgage	
Full-time	28	<\$40,000	11
Part-time	7	\$40,001–\$60,000	0
Homemaker	0	\$60,001–\$100,000	–9
Retired	15	\$100,000–\$150,000	–12
Student	–8	\$150,000–\$250,000	–16
Unemployed	–42	> \$250,000	–19
		Not a home owner	0
Time in current employment		Number of credit cards	
Not in full or part-time employment	0	0	–17
<1 year	–25	1–3	12
1–3 years	–10	4–8	0
1–10 years	0	9+	–18
> 10 year	31		
Residential status		Number of days currently past due on existing credit agreements	
Home owner	26	0 (All accounts up to date)	14
Renting	0	1–30 days past due	0
Living with parent	0	31–60 days past due	–29
		>60 days past due	–41
Loan amount requested as proportion of annual income		Declared bankrupt within the last 5 years?	
<10%	43	Yes	–50
10%–25%	22	No	9
26%–60%	0	Unknown	0
> 60%	–28		

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση; (παράδειγμα)

Για να υπολογίσετε το πιστωτικό αποτέλεσμά σας από το μοντέλο στο προηγούμενο σχήμα ξεκινάτε με το σταθερό σκορ 670.

Στη συνέχεια, περνάτε από την κάρτα βαθμολογίας ένα χαρακτηριστικό κάθε φορά, προσθέτοντας ή αφαιρώντας τα σημεία που ισχύουν για εσάς, έτσι εάν η κατάσταση απασχόλησής σας είναι πλήρους απασχόλησης, προσθέτετε 28 πόντους για να πάρετε 698.

Στη συνέχεια, αν σας ο χρόνος στην τρέχουσα απασχόληση είναι ας πούμε, δύο χρόνια, αφαιρείτε 10 πόντους για να πάρετε 688.

Εάν είστε ιδιοκτήτης σπιτιού, προσθέστε 26 πόντους.

Τι είναι η εξόρυξη δεδομένων και η προγνωστική ανάλυση; (παράδειγμα)

Η κατανομή βαθμολογίας δείχνει τη σχέση μεταξύ του πιστωτικού αποτελέσματος των ανθρώπων και των πιθανοτήτων αθέτησης.

Με σκορ 500 οι πιθανότητες είναι 1:1.

Αυτό σημαίνει ότι κατά μέσο όρο οι μισοί από αυτούς που βαθμολογούν 500 θα αθετήσουν εάν τους χορηγηθεί δάνειο.

Ομοίως, για όσους βαθμολογούν 620 οι πιθανότητες είναι 64:1.

Δηλαδή εάν παίρνετε 65 δανειολήπτες που βαθμολογούνται με 620, η προσδοκία είναι ότι το 64% θα επιστρέψει το δάνειο, αλλά στην πραγματικότητα δεν θα το κάνει.

Πόσο καλά είναι τα μοντέλα στην πρόβλεψη της συμπεριφοράς; (παράδειγμα)

Κατά μία έννοια, τα περισσότερα προγνωστικά μοντέλα είναι αρκετά φτωχά στην πρόβλεψη του τρόπου με τον οποίο κάποιος πρόκειται να συμπεριφερθεί.

Για να το δείξουμε αυτό, ας σκεφτούμε ένα παραδοσιακό πλάνο προώθησης πωλήσεων με βάση την αλληλογραφία.

Αν και σε παρακμή, οι αποστολή γραμμάτων παραμένει ένα δημοφιλές εργαλείο που χρησιμοποιούν οι επαγγελματίες του μάρκετινγκ για την προώθηση προϊόντων και υπηρεσιών στους καταναλωτές.

Πόσο καλά είναι τα μοντέλα στην πρόβλεψη της συμπεριφοράς; (παράδειγμα)

Για παράδειγμα, ασφαλιστική εταιρεία με μια στρατηγική μάρκετινγκ που περιλαμβάνει την αποστολή ενημερωτικών εντύπων σε πιθανούς πελάτες στα οποία έντυπα τους προσφέρουν μια πραγματικά καλή συμφωνία ασφάλισης ζωής.

Η εταιρεία έχει αποκτήσει μια βάση δεδομένων πιθανών δανειοληπτών από την οποία αντλεί δεδομένα προκειμένου να χρησιμοποιεί ένα μοντέλο πρόβλεψης για να προσδιορίσει ποιος από αυτούς είναι πιο πιθανό να θέλει ασφάλιση ζωής.

Με βάση τις προβλέψεις του μοντέλου γίνεται η αποστολή των ενημερωτικών εντύπων και συμβολαίων.

Πόσο καλά είναι τα μοντέλα στην πρόβλεψη της συμπεριφοράς; (παράδειγμα)

Εάν το μοντέλο είναι πραγματικά καλό, τότε η εταιρεία μπορεί να εντοπίσει άτομα με 1 στις 10 πιθανότητες να δεχτούν την προσφορά - 10 από κάθε 100 άτομα που ταχυδρομούνται ανταποκρίνονται.

Για να το θέσουμε με άλλο τρόπο, το μοντέλο θα είναι σωστό μόνο στο 10% του χρόνου και θα το κάνει λάθος στο 90% του χρόνου.

Προφανώς, αυτό είναι ένα αρκετά υψηλό ποσοστό αποτυχίας! Ωστόσο, αυτό που πρέπει να εξετάσετε είναι τι θα συνέβαινε χωρίς το μοντέλο.

Εάν επιλέξετε άτομα από τον τηλεφωνικό κατάλογο τυχαία, τότε ένα ποσοστό απόκρισης περίπου 1% είναι αρκετά χαρακτηριστικό για προωθήσεις πωλήσεων με αλληλογραφία.

Αν το κοιτάξετε με αυτόν τον τρόπο, τότε το μοντέλο είναι δέκα φορές καλύτερο από μια καθαρά τυχαία προσέγγιση - κάτι που δεν είναι καθόλου κακό.

Ποια είναι τα οφέλη των προγνωστικών μοντέλων;

Η παραδοσιακή προσέγγιση στη λήψη αποφάσεων βασίζεται στην γνώμη των πλέων έμπειρων επαγγελματιών.

Συνεχίζοντας με το παράδειγμα πιστοληπτικής βαθμολόγησης, δεν υπάρχει κανένας λόγος για τον οποίο οι Διευθυντές των τραπεζών δεν μπορούν να λάβουν αποφάσεις δανεισμού σχετικά με τους πελάτες τους (κάτι που έκαναν τις ημέρες πριν από την πιστωτική βαθμολόγηση).

Θα μπορούσε κανείς να υποστηρίξει ότι αυτό θα προσθέσει αυτή την προσωπική επαφή και ένας έμπειρος διευθυντής τράπεζας θα πρέπει να είναι σε καλύτερη θέση να αξιολογήσει την πιστοληπτική ικανότητα των πελατών.

Γιατί λοιπόν να χρησιμοποιήσετε μοντέλα πρόβλεψης;

Ποια είναι τα οφέλη των προγνωστικών μοντέλων;

Ένα όφελος είναι η ταχύτητα.

Όταν τα μοντέλα πρόβλεψης χρησιμοποιούνται ως μέρος ενός αυτοματοποιημένου συστήματος λήψης αποφάσεων, εκατομμύρια πελάτες μπορούν να αξιολογηθούν και να αντιμετωπιστούν μέσα σε λίγα δευτερόλεπτα.

Εάν μια τράπεζα θέλει να παράγει έναν κατάλογο πελατών πιστωτικών καρτών που μπορεί επίσης να είναι καλό για ένα δάνειο αυτοκινήτου, ένα μοντέλο πρόβλεψης επιτρέπει αυτό να γίνει γρήγορα και με σχεδόν μηδενικό κόστος.

Τέτοια συστήματα επιτρέπουν τη λήψη αποφάσεων σε πραγματικό χρόνο ενώ ο πελάτης είναι στο τηλέφωνο, στο υποκατάστημα ή στο διαδίκτυο.

Ποια είναι τα οφέλη των προγνωστικών μοντέλων;

Ένα δεύτερο σημαντικό όφελος από τη χρήση προγνωστικών μοντέλων είναι ότι γενικά **κάνουν καλύτερες προβλέψεις από τους ανθρώπους**. Πόσο καλύτερες εξαρτάται από το συγκεκριμένο πρόβλημα να είναι δύσκολο να ποσοτικοποιηθούν.

Ωστόσο, από την εμπειρία μου, θα περίμενα ένα καλά εφαρμοσμένο σύστημα λήψης αποφάσεων, βασισμένο σε προγνωστικά αναλυτικά στοιχεία, να λαμβάνει αποφάσεις που είναι περίπου **20-30%** πιο ακριβείς από τους πλέον έμπειρους ανθρώπους.

Ποια είναι τα οφέλη των προγνωστικών μοντέλων;

Στο παράδειγμα βαθμολόγησης πίστωσης, αυτό μεταφράζεται στη χορήγηση 20-30% λιγότερων Δανείων σε πελάτες που θα είχαν αθετήσει ή 20-30% περισσότερα δάνεια σε καλούς πελάτες που θα αποπληρώσουν, ανάλογα με το πώς κάποιος αποφασίζει να χρησιμοποιήσει το μοντέλο.

Για να το θέσω διαφορετικά, αν μια τράπεζα έχει σε ετήσια βάση δάνεια ανεπίδεκτα εισπράξεως της τάξεως των €500 εκατ., μια λογική προσδοκία είναι ότι αυτό θα μπορούσε να μειωθεί κατά τουλάχιστον €100 εκατ., αν όχι περισσότερο, με τη χρήση predictive analytics

Ποια είναι τα οφέλη των προγνωστικών μοντέλων;

Αν μιλάμε για ένα τμήμα μάρκετινγκ που δαπανά 20 εκατομμύρια Ευρώ για άμεσο μάρκετινγκ για να προσελκύσει 300.000 νέους πελάτες κάθε χρόνο, τότε υιοθετώντας προγνωστικά αναλυτικά στοιχεία θα περίμενε κανείς να δαπανήσει περίπου 5 εκατομμύρια Ευρώ λιγότερο για να προσελκύσει τον ίδιο αριθμό πελατών.

Εναλλακτικά, θα μπορούσαν να αναμένουν να προσελκύσουν περίπου 75.000 περισσότερους πελάτες για την ίδια δαπάνη 20 εκατομμυρίων Ευρώ.

Ποια είναι τα οφέλη των προγνωστικών μοντέλων;

Ένα τρίτο όφελος είναι η **συνέπεια**.

Ένα δεδομένο μοντέλο πρόβλεψης θα παράγει πάντα την ίδια πρόβλεψη όταν παρουσιάζεται με τα ίδια δεδομένα.

Αυτό δεν συμβαίνει με τους ανθρώπινους υπεύθυνους λήψης αποφάσεων.

Υπάρχουν πολλές αποδείξεις ότι ακόμη και ο πιο ικανός εμπειρογνώμονας θα καταλήξει σε πολύ διαφορετικά συμπεράσματα και θα λάβει διαφορετική απόφαση για κάτι ανάλογο με τη διάθεσή τους, την ώρα της ημέρας, αν είναι πεινασμένοι ή όχι και πλήθος άλλων παραγόντων.

Ποια είναι τα οφέλη των προγνωστικών μοντέλων;

Αυτό οδηγεί σε ερωτήσεις σχετικά με την προκατάληψη που μερικοί άνθρωποι εμφανίζουν (συνειδητά ή ασυνείδητα) εναντίον ανθρώπων λόγω του φύλου, της φυλής, της θρησκευτικής ηλικίας, του σεξουαλικού προσανατολισμού και ούτω καθεξής.

Αυτό δεν σημαίνει ότι τα προγνωστικά μοντέλα δεν εμφανίζουν προκατάληψη προς μία ή την άλλη ομάδα, αλλά ότι όπου υπάρχει προκατάληψη βασίζεται σε σαφή στατιστικά στοιχεία

Ποια είναι τα οφέλη των προγνωστικών μοντέλων;

Πολλοί τύποι προγνωστικού μοντέλου, όπως η κάρτα βαθμολογίας, εξηγούνται. Είναι εύκολο να καταλάβετε πώς κάποιος πήρε το σκορ που έκαναν και, ως εκ τούτου, γιατί έκαναν ή δεν έλαβαν δάνειο.

Η εξήγηση του γιατί ένας εμπειρογνώμονας πήρε μια συγκεκριμένη απόφαση δεν είναι πάντα τόσο εύκολη, ειδικά αν βασίστηκε σε ένα προαίσθημα.

Ακόμη και αν ο υπεύθυνος λήψης αποφάσεων διατηρεί λεπτομερείς σημειώσεις, Η ερμηνεία του τι εννοούσαν δεν είναι πάντα εύκολη μετά το συμβάν.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Η βαθμολόγηση πιστώσεων ήταν η πρώτη εμπορική εφαρμογή των προγνωστικών αναλυτικών στοιχείων (και παραμένει μία από τις πιο δημοφιλείς) και από τη δεκαετία του 1980 οι ίδιες μέθοδοι εφαρμόστηκαν σε άλλους τομείς των χρηματοπιστωτικών υπηρεσιών.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Στα τμήματα μάρκετινγκ τους, οι πάροχοι δανείων και πιστωτικών καρτών άρχισαν να αναπτύσσουν μοντέλα για να προσδιορίσουν την πιθανότητα ανταπόκρισης σε μια επικοινωνία μάρκετινγκ, έτσι ώστε μόνο όσοι είναι πιθανότερο να ενδιαφέρονται για ένα προϊόν να στοχεύουν με μια προσφορά.

Αυτό γλίτωσε τεράστια ποσά σε σύγκριση με τις γενικές στρατηγικές μάρκετινγκ που προηγήθηκαν και επέτρεψε την αποστολή εξατομικευμένων επικοινωνιών σε κάθε άτομο με βάση το σκορ που έλαβαν.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Ομοίως, άρχισαν να χρησιμοποιούνται στα ασφαλιστικά μοντέλα για την πρόβλεψη της πιθανότητας και της αξίας των απαιτήσεων.

Αυτές οι προβλέψεις χρησιμοποιήθηκαν στη συνέχεια για τον καθορισμό ασφαλίστρων.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Στις μέρες μας, τα προγνωστικά μοντέλα χρησιμοποιούνται σχεδόν για κάθε είδος πρόβλημα που μπορεί να αντιμετωπίσει μια επιχείρηση και όχι μόνο.

- Αναγνώριση ατόμων που δεν πληρώνουν τους φόρους τους.
- Υπολογισμός της πιθανότητας εμφάνισης εγκεφαλικού επεισοδίου τα επόμενα 10 χρόνια.
- Εντοπίζοντας ποιες συναλλαγές με πιστωτικές κάρτες είναι δόλιες.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

- Επιλογή υπόπτων σε ποινικές υποθέσεις.
- Η απόφαση πρόσληψης προσωπικού.
- Προβλέποντας πόσο πιθανό είναι ότι ένας πελάτης θα χρεοκοπήσει.
- Καθορίζοντας ποια βιβλία, μουσική και ταινίες είναι πιθανό να αγοράσετε στη συνέχεια.
- Προβλέποντας πόσο είναι πιθανό να περάσετε από το τοπικό σας σούπερ μάρκετ την επόμενη εβδομάδα.
- Πρόβλεψη προσδόκιμου ζωής.
- Εκτίμηση του ποσού που θα δαπανήσει κάποιος για την πιστωτική τους κάρτα φέτος.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Οι εφαρμογές των προγνωστικών μοντέλων στην παραπάνω λίστα εμπίπτουν σε δύο ομάδες.

Εκείνοι στην πρώτη ομάδα ασχολούνται με ερωτήσεις τύπου **Ναι/Όχι** σχετικά με τη συμπεριφορά.

Θα κάνει κάποιος κάτι ή όχι;

Θα πραγματοποιήσουν τη δράση A ή τη δράση B;

Τα μοντέλα που προβλέπουν αυτό το είδος συμπεριφοράς ονομάζονται **μοντέλα ταξινόμησης**.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Η έξοδος αυτών των μοντέλων (η βαθμολογία μοντέλου) είναι ένας αριθμός που αντιπροσωπεύει την πιθανότητα (οι πιθανότητες) της συμπεριφοράς.

Μερικές φορές η βαθμολογία παρέχει μια άμεση εκτίμηση της πιθανότητας συμπεριφοράς.

Για παράδειγμα, μια βαθμολογία 0,4 σημαίνει ότι η πιθανότητα κάποιος να έχει καρδιακή προσβολή τα επόμενα πέντε χρόνια είναι 40% (και ως εκ τούτου υπάρχει 60% πιθανότητα να μην έχουν).

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Σε άλλες περιπτώσεις η βαθμολογία βαθμονομείται σε μια δεδομένη κλίμακα - ίσως 100 σημαίνει ότι η πιθανότητα να έχετε καρδιακή προσβολή είναι η ίδια με τον μέσο όρο του πληθυσμού.

Μια βαθμολογία 200 δύο φορές κατά μέσο όρο, μια βαθμολογία 400 τέσσερις φορές κατά μέσο όρο και ούτω καθεξής.

Για την κάρτα βαθμολογίας που είδαμε, οι πιθανότητες αθέτησης διπλασιάζονται κάθε 20 πόντους - κάτι που είναι παρόμοια κλίμακα με αυτό που χρησιμοποιεί η FICO στα πιστωτικά της αποτελέσματα.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Ένα ηλεκτρονικό βιβλιοπωλείο μπορεί να χτίσει ένα μοντέλο αναλύοντας το κείμενο σε βιβλία που οι άνθρωποι έχουν αγοράσει στο παρελθόν για να προβλέψουν τα βιβλία που αγόρασαν στη συνέχεια.

Μόλις δημιουργηθεί αυτό το μοντέλο, τότε το παρελθόν ιστορικό αγορών σας μπορεί να χρησιμοποιηθεί μέσω του μοντέλου για να δημιουργήσει ένα σκορ για κάθε βιβλίο στη λίστα του βιβλιοπώλη.

Όσο υψηλότερη είναι η βαθμολογία, τόσο πιο πιθανό είναι να αγοράσετε κάθε βιβλίο. Ο έμπορος στη συνέχεια σας εμφανίζει τα δύο ή τρία βιβλία που βαθμολογούν περισσότερο: **αυτά που είναι πιο πιθανό να ενδιαφέρονται να αγοράσουν.**

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Ο δεύτερος τύπος προγνωστικού μοντέλου αφορά τις ποσότητες.

Δεν πρόκειται για το αν πρόκειται να κάνετε κάτι ή όχι, αλλά για το μέγεθος αυτού που κάνετε.

Συνήθως, αυτά ισοδυναμούν με ερωτήσεις τύπου " πόσο " ή " πόσο καιρό ". Οι αναλογιστές χρησιμοποιούν προγνωστικά μοντέλα για να προβλέψουν πόσο καιρό θα ζήσουν οι άνθρωποι και επομένως τι είδους σύνταξη μπορούν να περιμένουν.

Οι εταιρείες πιστωτικών καρτών κατασκευάζουν μοντέλα αξίας για να εκτιμήσουν πόσα έσοδα είναι πιθανό να δημιουργήσει κάθε πελάτης.

Αυτοί οι τύποι μοντέλων ονομάζονται μοντέλα παλινδρόμησης

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Συνήθως, η βαθμολογία από ένα μοντέλο παλινδρόμησης παρέχει μια άμεση εκτίμηση της ποσότητας ενδιαφέροντος.

Μια βαθμολογία 1.500 που παράγεται από ένα μοντέλο εσόδων σημαίνει ότι ο πελάτης αναμένεται να δαπανήσει 1.500 Ευρώ.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Ωστόσο, μερικές φορές αυτό που ενδιαφέρει κάποιον είναι η κατάταξη των πελατών, παρά οι απόλυτες τιμές.

Το μοντέλο μπορεί να κατασκευαστεί για να παράγει βαθμολογίες στην περιοχή 1-100, που αντιπροσωπεύουν το εκατοστημόριο στο οποίο πέφτουν οι δαπάνες των πελατών.

Ένα σκορ 1 δείχνει ότι ο πελάτης βρίσκεται στο χαμηλότερο εκατοστημόριο δαπανών και ένα σκορ 100 ότι βρίσκεται στο υψηλότερο εκατοστημόριο βαθμολόγησης.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Όσον αφορά τον τρόπο εμφάνισης, τα μοντέλα ταξινόμησης και παλινδρόμησης είναι πολύ παρόμοια, αλλά σε τεχνικό επίπεδο υπάρχουν λεπτές διαφορές που καθορίζουν τον τρόπο κατασκευής και χρήσης των μοντέλων.

Τα μοντέλα ταξινόμησης εφαρμόζονται ευρύτερα, αλλά τα μοντέλα παλινδρόμησης είναι όλο και πιο δημοφιλή επειδή δίνουν μια πολύ πιο λεπτομερή εικόνα της συμπεριφοράς των πελατών.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Παλαιότερα ένα ενιαίο μοντέλο πιστωτικής βαθμολόγησης θα είχε χρησιμοποιηθεί για να προβλέψει εάν ή όχι κάποιος ήταν πιθανό να εξοφλήσει το δάνειο του, αλλά στις μέρες μας οι δανειστές δημιουργούν επίσης μοντέλα για να προβλέψουν την αναμενόμενη ζημία για την αθέτηση δανείων και τα αναμενόμενα έσοδα από πελάτες που αποπληρώνουν τα δάνειά τους.

Κρατώντας τα οφέλη & αποφεύγοντας τις παγίδες

Ένας οργανισμός που εφαρμόζει καλά τα προγνωστικά αναλυτικά στοιχεία μπορεί να αναμένει να δει βελτιώσεις στις επιχειρηματικές του διαδικασίες κατά 20-30% ή και περισσότερο σε ορισμένες περιπτώσεις.

Ωστόσο, η επιτυχία δεν είναι καθόλου εγγυημένη.

Κρατώντας τα οφέλη & αποφεύγοντας τις παγίδες

Στην πρώτη μου δουλειά, που εργαζόμουν για μια εταιρεία αξιολόγησης πιστώσεων, ασχολήθηκα με την δημιουργία μοντέλων πρόβλεψης για διάφορους πελάτες.

Σε γενικές γραμμές τα πράγματα πήγαν αρκετά καλά. Παρέδωσα μοντέλα πρόβλεψης καλής ποιότητας και οι πελάτες μας ήταν ευχαριστημένοι και οι απολαβές που είχαμε ήταν ανάλογες.

Ωστόσο, σε μεταγενέστερη επικοινωνία με τους πελάτες μας διαπίστωσα ότι δεν ήταν όλοι ευχαριστημένοι.

Πολλά από τα μοντέλα που είχα αναπτύξει είχαν εφαρμοστεί και παρείχαν πραγματικά οφέλη, αλλά αυτό δεν συνέβαινε καθολικά. Ορισμένα μοντέλα δεν είχαν εφαρμοστεί ή η εφαρμογή είχε αποτύχει για κάποιο λόγο.

Κρατώντας τα οφέλη & αποφεύγοντας τις παγίδες

Αναλύοντας το πρόβλημα λίγο καλύτερα έγινε φανερό ότι δεν ήταν πρόβλημα των μοντέλων, αλλά ήταν μια σειρά από οργανωτικά και πολιτιστικά ζητήματα που ήταν το πρόβλημα.

Υπάρχουν πολλοί λόγοι για τους οποίους ένα έργο προγνωστικής ανάλυσης μπορεί να αποτύχει, αλλά αυτά μπορούν συνήθως να τοποθετηθούν σε μία από τις τρεις κατηγορίες:

Κρατώντας τα οφέλη & αποφεύγοντας τις παγίδες

Η εταιρεία δεν είναι έτοιμη για προγνωστική ανάλυση.

Κάνοντας κάτι νέο είναι επικίνδυνο. Οι άνθρωποι είναι συχνά απρόθυμοι να πάρουν το ρίσκο που απαιτείται για να δώσουν εμπιστοσύνη στα αυτοματοποιημένα μοντέλα και όχι στην ανθρώπινη κρίση.

Το λάθος μοντέλο.

Ο κατασκευαστής μοντέλων σκέφτηκε ότι ο πελάτης τους ήθελε ένα μοντέλο για να προβλέψει έναν τύπο συμπεριφοράς των καταναλωτών, αλλά ο πελάτης ήθελε πραγματικά κάτι που προέβλεπε μια διαφορετική συμπεριφορά.

Κρατώντας τα οφέλη & αποφεύγοντας τις παγίδες

Αδύναμη διακυβέρνηση.

Η εφαρμογή ενός προγνωστικού μοντέλου απαιτεί μερικές φορές αλλαγές στις πρακτικές εργασίας. Κατά κανόνα, οι άνθρωποι δεν τους αρέσει η αλλαγή και δεν θα αλλάξουν αν δεν πρέπει. Απλά λέγοντάς τους να κάνουν κάτι διαφορετικό ή να εκδώσουν μερικά σημειώματα δεν λειτουργεί. **Χρειάζεται αποτελεσματική διαχείριση και η επιβολή είναι απαραίτητη.**

Κρατώντας τα οφέλη & αποφεύγοντας τις παγίδες

Ένα πράγμα που πρέπει να έχουμε κατά νου είναι ότι διαφορετικοί άνθρωποι έχουν διαφορετικές απόψεις για το ποιες είναι οι συνέπειες ενός έργου.

Για έναν επιστήμονα δεδομένων, ένα πρόγραμμα προγνωστικής ανάλυσης αφορά τη συλλογή δεδομένων και στη συνέχεια την οικοδόμηση του καλύτερου (πιο προγνωστικού) μοντέλου.

Ο τρόπος εφαρμογής και η χρησιμότητα του μοντέλου δεν του προκαλεί μεγάλη ανησυχία. Τα ευρύτερα ζητήματα γύρω από την εφαρμογή, τις οργανωτικές δομές και τον πολιτισμό είναι πολύ έξω από το τι θέλουν να πετύχουν με το μοντέλο.

Τι είναι τα Big Data;

Μεγάλα και σύνθετα σύνολα δεδομένων υπάρχουν εδώ και δεκαετίες.

Από μια άποψη, τα μεγάλα δεδομένα δεν είναι κάτι καινούργιο, και για μερικούς στη βιομηχανία όλη η διαφημιστική εκστρατεία γύρω από τα μεγάλα δεδομένα ήρθε ως ένα κομμάτι μιας έκπληξης.

Από τις αρχές της δεκαετίας του 2010, τα "μεγάλα δεδομένα" είχαν γίνει η δημοφιλής φράση catch-all για να περιγράψουν βάσεις δεδομένων που δεν είναι μόνο μεγάλες, αλλά τεράστιες και πολύπλοκες. Δεν υπάρχει καθολικά συμφωνημένος ορισμός των "μεγάλων δεδομένων", αλλά τα χαρακτηριστικά των μεγάλων δεδομένων που θεωρούνται σημαντικά είναι:

Τι είναι τα Big Data;

Όγκος (volume). Οποιαδήποτε βάση δεδομένων που είναι πολύ μεγάλη για να διαχειρίζεται άνετα σε έναν μέσο υπολογιστή/φορητό υπολογιστή/διακομιστή μπορεί να θεωρηθεί μεγάλα δεδομένα.

Κατά τη στιγμή της γραφής, τα μεγάλα δεδομένα θεωρούνται γενικά ως βάση δεδομένων που περιέχει περισσότερα από ένα terabyte (1.000 gigabytes) δεδομένων. Ορισμένες μεγάλες πηγές δεδομένων περιέχουν petabytes δεδομένων (1 petabyte = 1.000 terabytes).

Τι είναι τα Big Data;

Ποικιλία (Variety). Τα μεγάλα δεδομένα περιέχουν πολλούς διαφορετικούς τύπους δομημένων και αδόμητων δεδομένων.

Τα δομημένα δεδομένα είναι τακτοποιημένα και καλά καθορισμένα και συνήθως μπορούν να αναπαρασταθούν ως αριθμοί ή κατηγορίες: για παράδειγμα το εισόδημά σας, η ηλικία, το φύλο και η οικογενειακή σας κατάσταση.

Τα μη δομημένα δεδομένα δεν είναι καλά καθορισμένα. Είναι συχνά κείμενο και δύσκολο να κατηγοριοποιηθεί: για παράδειγμα ηλεκτρονικά μηνύματα, ιστολόγια, ιστοσελίδες και μεταγραφές τηλεφωνικών συνομιλιών.

Τι είναι τα Big Data;

Μεταβλητότητα (ταχύτητα) [Volatility (velocity)]. Ορισμένοι τύποι δεδομένων είναι σχετικά στατικοί, όπως ο τόπος γέννησης, το φύλο, ο αριθμός κοινωνικής ασφάλισης και η εθνικότητα κάποιου.

Άλλα δεδομένα αλλάζουν περιστασιακά / αργά, όπως η διεύθυνση κάποιου, ο εργοδότης σας ή ο αριθμός των παιδιών που έχετε.

Στο άλλο άκρο τα δεδομένα αλλάζουν συνεχώς: για παράδειγμα ποια μουσική ακούτε τώρα, την ταχύτητα που οδηγείτε και τον καρδιακό σας ρυθμό. **Τα μεγάλα δεδομένα είναι συχνά ασταθή.**

Τι είναι τα Big Data;

Πολλαπλών πηγών (Multi-sourced). Ορισμένες πηγές Big Data δημιουργούνται εξ ολοκλήρου από τα εσωτερικά συστήματα ενός οργανισμού. Αυτό σημαίνει ότι έχουν τον έλεγχο της δομής και της μορφής του.

Ωστόσο, τα Big Data περιλαμβάνει συχνά εξωτερικά δεδομένα, όπως πιστωτικές αναφορές, πληροφορίες απογραφής, δεδομένα GPS και ιστοσελίδες και οι οργανισμοί έχουν ελάχιστο έλεγχο σχετικά με τον τρόπο παροχής και μορφοποίησης.

Αυτό εισάγει πρόσθετα ζητήματα σχετικά με την ποιότητα των δεδομένων, το απόρρητο και την ασφάλεια, πέρα από αυτό που απαιτείται από δεδομένα εσωτερικής προέλευσης.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Προληπτικά να αναζητάτε και να αποκτάτε νέα δεδομένα: συγκεντρώνετε όλα τα δεδομένα σας και τα αναλύετε για να παράγετε πληροφορίες σχετικά με το τι έχουν κάνει οι άνθρωποι, τι κάνουν και τι είναι πιθανό να κάνουν στο μέλλον. Αυτό με τη σειρά του πληροφορεί για τη λήψη αποφάσεων και ποιες ενέργειες πρέπει να κάνετε.

Μια μεγάλη φιλοσοφία δεδομένων είναι να έχετε μια ολιστική άποψη των δεδομένων που έχετε στη διάθεσή σας και να πάρετε το καλύτερο από αυτό που έχετε.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Εάν ένας οργανισμός το κάνει αυτό τότε δεν έχει σημασία αν έχει μόνο λίγα megabytes ή πολλά petabytes δεδομένων. εάν τα δεδομένα είναι δομημένα ή αδόμητα ή από πού προέρχονται.

Από την άποψη της τεχνολογίας θέλουμε λύσεις που δίνουν την απαιτούμενη αποθήκευση και αναλυτική ικανότητα. Σε ορισμένες περιπτώσεις αυτό μπορεί να σημαίνει τη χρήση νεότερων τεχνολογιών όπως το Hadoop ή το Storm, αλλά μια παραδοσιακή λύση σχεσιακής βάσης δεδομένων είναι συχνά επαρκής (ή ακόμα και ανώτερη) και δεν πρέπει να αποκλειστεί.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Έτσι, μπορείτε να δείτε τα μεγάλα δεδομένα από διάφορες οπτικές γωνίες, αλλά για το υπόλοιπο του μαθήματος του βιβλίου θα κρατήσουμε τα πράγματα απλά και θα υιοθετήσουμε έναν αρκετά χαλαρό ορισμό των μεγάλων δεδομένων ως: **"ένα πολύ μεγάλο όγκο ποικίλων δεδομένων."**

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Μια κυβερνητική απογραφή που περιέχει μερικές δεκάδες πληροφορίες για κάθε ένα από τα 250 εκατομμύρια πολίτες της χώρας, όπως ο αριθμός κοινωνικής ασφάλισης, το φύλο, η οικογενειακή κατάσταση, το εισόδημα και ο αριθμός των παιδιών, είναι ένας μεγάλος όγκος δεδομένων, αλλά πιθανώς όχι μεγάλα δεδομένα (big data).

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Ομοίως, το κείμενο από 50.000.000 σελίδες στο Διαδίκτυο δεν είναι ούτε μεγάλα δεδομένα.

Ωστόσο, η βάση δεδομένων ενός οργανισμού με όλες τις πληροφορίες που έχει συγκεντρώσει για τα τρία εκατομμύρια πελάτες του τα τελευταία πέντε χρόνια (λεπτομέρειες αγοράς, πληροφορίες χρέωσης, ηλεκτρονικά μηνύματα, αρχεία καταγραφής διακομιστών, κείμενα, μεταγραφές τηλεφωνικών κλήσεων, επιστολές παραπόνων, σημειώσεις που λαμβάνονται από το προσωπικό στο υποκατάστημα, πιστωτικές εκθέσεις, δεδομένα GPS κ. ο.Κ.) είναι μεγάλα δεδομένα.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Ο πρώτος μύθος που θέλω να καταρύψουμε είναι ότι χρειάζεστε ένα τεράστιο όγκο δεδομένων για να δημιουργήσετε ένα προγνωστικό μοντέλο. Μερικές χιλιάδες αρχεία πελατών και μερικές δεκάδες πληροφορίες σχετικά με αυτούς τους πελάτες είναι περισσότερο από αρκετές και πολλά χρήσιμα μοντέλα έχουν κατασκευαστεί χρησιμοποιώντας λιγότερα δεδομένα από αυτό.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Ωστόσο, όσο περισσότερα δεδομένα έχετε για τους ανθρώπους, τόσο πιο προγνωστικά θα είναι τα μοντέλα σας.

Τα μεγάλα δεδομένα έχουν προσελκύσει τόσο μεγάλο ενδιαφέρον τα τελευταία χρόνια, διότι υπερβαίνουν τις παραδοσιακές πηγές δεδομένων που έχουν χρησιμοποιήσει οι άνθρωποι στην εξόρυξη δεδομένων και στην προγνωστική ανάλυση στο παρελθόν.

Συγκεκριμένα, όταν οι άνθρωποι μιλούν για μεγάλα δεδομένα, αυτό που συχνά σημαίνουν είναι:

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Δεδομένα κειμένου. Αυτό προέρχεται από γράμματα, μεταγραφές τηλεφώνου, e-mail, ιστοσελίδες, tweets και ούτω καθεξής. Αυτός ο τύπος δεδομένων δεν είναι δομημένος και επομένως χρειάζεται πολλή δύναμη επεξεργασίας για να τα αναλύσει.

Δεδομένα που δημιουργούνται από μηχανή. Για παράδειγμα, δεδομένα GPS από τηλέφωνα ατόμων, αρχεία καταγραφής ιστού που παρακολουθούν τη χρήση διαδικτύου και τηλεματικές συσκευές που είναι τοποθετημένες σε αυτοκίνητα. Τα δεδομένα που δημιουργούνται από μηχανήματα είναι γενικά καλά δομημένα και εύκολα στην ανάλυση, αλλά είναι πάρα πολλά σε όγκο.

Εφαρμογές προγνωστικών αναλυτικών μοντέλων

Δεδομένα δικτύου. Αυτές είναι πληροφορίες για την οικογένεια των ανθρώπων, τους φίλους και άλλους συνεργάτες.

Σε ορισμένα πλαίσια, αυτό που είναι σημαντικό είναι η δομή του δικτύου στο οποίο ανήκει ένα άτομο - πόσα άτομα βρίσκονται στο δίκτυο, που βρίσκονται στο κέντρο του δικτύου και ούτω καθεξής.

Σε άλλα πλαίσια, το δίκτυο είναι ένας μηχανισμός για την εξαγωγή πραγμάτων για κάποιον, βάσει των χαρακτηριστικών άλλων ατόμων στο δίκτυό του.

Πόση αξία προσθέτουν τα Big Data;

Οι μακροχρόνιοι χρήστες προγνωστικών αναλυτικών στοιχείων, όπως οι τράπεζες, οι έμποροι και οι λιανοπωλητές, ρωτούν για την αξία των Μεγάλων Δεδομένων, αυτό που πραγματικά θέλουν να μάθουν είναι: ποια είναι η βελτίωση στα προγνωστικά μας μοντέλα που προέρχεται από τα μεγάλα δεδομένα, πέρα από τα δεδομένα που χρησιμοποιούμε ήδη;

Ένα χαρακτηριστικό των Μεγάλων Δεδομένων είναι ότι το μεγαλύτερο μέρος του έχει πολύ χαμηλή πυκνότητα πληροφοριών, καθιστώντας πολύ δύσκολη την εξαγωγή χρήσιμων πληροφοριών από τον πελάτη.

Πόση αξία προσθέτουν τα Big Data;

Ένα τεράστιο ποσοστό των Μεγάλων Δεδομένων είναι απολύτως άχρηστο όταν πρόκειται για την πρόβλεψη της συμπεριφοράς των καταναλωτών.

Πρέπει να εργαστείτε πολύ σκληρά για να βρείτε τα χρήσιμα κομμάτια που θα βελτιώσουν την ακρίβεια των προγνωστικών μοντέλων σας - και γι ' αυτό χρειάζεστε μεγάλους υπολογιστές με μεγάλη αποθηκευτική ικανότητα και έξυπνους αλγόριθμους για να βρείτε τα σημαντικά πράγματα ανάμεσα στα μη-σημαντικά.

Πόση αξία προσθέτουν τα Big Data;

Ο τρόπος που μου αρέσει να σκέφτομαι αυτό το θέμα - όπως και πολλοί άλλοι - είναι κατ' αναλογία με την εξόρυξη χρυσού.

Η περισσότερη βιομηχανική δραστηριότητα που εμπλέκεται στην εξόρυξη χρυσού συμβαίνει σε ορυχεία όπου υπάρχουν πλούσια κοιτάσματα.

Ωστόσο, έχει εκτιμηθεί ότι υπάρχουν χιλιάδες τόνοι χρυσού που διαλύονται στους ωκεανούς του κόσμου, πιθανώς περισσότερο από το σύνολο του ποσού που έχει εξορυχθεί σε όλη την ανθρώπινη ιστορία.

Ωστόσο, ο χρυσός είναι πολύ διάχυτος στην θάλασσα και κανείς δεν έχει βρει ακόμα έναν οικονομικά αποδοτικό τρόπο εξαγωγής του.

Πόση αξία προσθέτουν τα Big Data;

Οι παραδοσιακές βάσεις δεδομένων πελατών είναι ανάλογες με το χρυσωρυχείο, ενώ τα μεγάλα δεδομένα είναι ο ωκεανός.

Υπάρχουν πολλές χρήσιμες πληροφορίες σε μεγάλα δεδομένα, αλλά είναι πολύ αραιές σε σύγκριση με τις παραδοσιακές πηγές δεδομένων και μπορεί να είναι πολύ ακριβό να φτάσουμε σε αυτά.

Πόση αξία προσθέτουν τα Big Data;

Μερικές φορές τα οφέλη των Big Data μπορεί να είναι πολύ μικρά για να δικαιολογήσουν τη δαπάνη.

Στον τραπεζικό τομέα, για παράδειγμα, το δυναμικό για νέες μεγάλες πηγές δεδομένων για τη βελτίωση της προγνωστικής ικανότητας των μοντέλων βαθμολόγησης πιστώσεων είναι αρκετά μικρό, πέρα από τα ήδη διαθέσιμα δεδομένα.

Αυτό οφείλεται στο γεγονός ότι ο βασικός μοχλός του πιστωτικού κινδύνου είναι η συμπεριφορά του παρελθόντος και οι τράπεζες έχουν έτοιμη πρόσβαση στις πιστωτικές εκθέσεις των ανθρώπων, καθώς και μια πληθώρα άλλων δεδομένων, που παρέχονται σε μια ωραία τακτοποιημένη μορφή από οργανισμούς αναφοράς πιστώσεων όπως Equifax, Experian και TransUnion.

Πόση αξία προσθέτουν τα Big Data;

Από την άλλη πλευρά, εάν είστε μια ομάδα μάρκετινγκ που προσπαθεί να εντοπίσει άτομα που μπορεί να ενδιαφέρονται για τα προϊόντα σας και δεν έχετε τίποτα να συνεχίσετε, τότε τα εξωτερικά δημιουργημένα μεγάλα δεδομένα μπορούν να προσφέρουν μεγάλη αξία.

Με το “ψάξετε” στο διαδίκτυο, αρχεία καταγραφής διακομιστών, ιστότοπους κοινωνικής δικτύωσης, tweets και ούτω καθεξής για να μάθετε ποιος μιλάει για το τι, ή τι αγοράζουν οι φίλοι και η οικογένεια των ανθρώπων έχει τεράστια αξία.