

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ**

ΕΡΓΑΣΙΑ ΣΤΟ ΜΑΘΗΜΑ «ΠΙΣΤΩΤΙΚΣ ΚΙΝΔΥΝΟΣ»

Μια τράπεζα ενδιαφέρεται να διαμορφώσει μοντέλα βαθμολόγησης πιστοληπτικής ικανότητας των πελατών της με βάση τα στοιχεία που έχει συγκεντρώσει για τη συμπεριφορά των δανειοληπτών της στον κλάδο της Λιανικής Τραπεζικής μέχρι σήμερα. Τα δεδομένα βρίσκονται στο αρχείο **data.xlsx** και αφορούν ένα δείγμα 3400 τυχαία επιλεγμένων δανειοληπτών για τους οποίους έχουν καταγραφεί οι ακόλουθες μεταβλητές:

Μεταβλητή	Περιγραφή
Status	Κατάσταση Δανείου (1: Μη εξυπηρετούμενο, 0: Εξυπηρετούμενο)
Loss	Ζημιά που προκλήθηκε λόγω της χρεοκοπίας του δανειολήπτη σε χιλιάδες Ευρώ
Cards	Αριθμός πιστωτικών καρτών που έχει ο πελάτης στην κατοχή του
Deposit	Καταθέσεις τη στιγμή της δανειοδότησης, σε χιλιάδες Ευρώ
Duration	Διάρκεια αποπληρωμής δάνειου
Family	Οικογενειακή κατάσταση (1: Έγγαμος 0: Άγαμος/Διαζευγμένος)
Score	Συνολική βαθμολογία σε ερωτηματολόγιο που συμπληρώθηκε από τον αιτούντα
Type	Είδος εργασίας (1:εποχική, 2: αορίστου χρόνου 3: με σύμβαση ορισμένου χρόνου)
Years	Χρόνια εργασίας

Αφού εξαιρεθούν οι γραμμές με αρίθμηση $5(i + 1)$ έως $5(i + 10)$, όπου i είναι οι δύο τελευταίοι αριθμοί του αριθμού μητρώου σας, να γίνει η επεξεργασία που περιγράφεται παρακάτω και να απαντηθούν τα αντίστοιχα ερωτήματα (όπου απαιτούνται στατιστικοί έλεγχοι, να γίνουν σε επίπεδο σημαντικότητας 5%).

- A. Να υπολογισθούν οι μέσες τιμές, οι τυπικές αποκλίσεις, οι μέγιστες και οι ελάχιστες τιμές των ποσοτικών χαρακτηριστικών για το σύνολο των δεδομένων. Να κατασκευασθούν επίσης πίνακες διπλής εισόδου(με συχνότητες και ποσοστά ως προς το σύνολο των παρατηρήσεων) για όλα τα ζεύγη κατηγορικών μεταβλητών. [Μονάδες 5]
- B. Χρησιμοποιώντας τη μέθοδο Forward:LR της Λογιστικής παλινδρόμησης να βρεθεί το καλύτερο μοντέλο για την πρόβλεψη της πιθανότητας να είναι μη εξυπηρετούμενο ένα δάνειο. Ως προσβλέπουσες μεταβλητές να χρησιμοποιηθούν όλες οι μεταβλητές του παραπάνω πίνακα με εξαίρεση τη μεταβλητή Loss. Ως όριο για την εισαγωγή μεταβλητής στο μοντέλο να χρησιμοποιηθεί το 5% και για την εξαγωγή το 10%. [Μονάδες 25]
 1. Ποιες μεταβλητές είναι στατιστικά σημαντικές και ποιες δεν χρειάζονται στο τελικό προτεινόμενο μοντέλο; Να δοθεί ο τύπος υπολογισμού της πιθανότητας να είναι ένα δάνειο εξυπηρετούμενο.
 2. Να δοθεί ο πίνακας ορθών ταξινομήσεων του μοντέλου που διαμορφώσατε. Πως θα μπορούσατε να αυξήσετε το ποσοστό των ορθών ταξινομήσεων των κακών πελατών ώστε αυτό να ξεπεράσει την τιμή 85%; Να δοθεί ο τύπος υπολογισμού της πιθανότητας να είναι ένα δάνειο εξυπηρετούμενο με το νέο μοντέλο.
 3. Να δοθεί η ποσοστιαία μεταβολή της σχετικής πιθανότητας χρεοκοπίας με διάστημα εμπιστοσύνης συντελεστή 95% στις ακόλουθες περιπτώσεις:
 - α. αν αυξηθεί ο αριθμός των καρτών ενός πελάτη κατά μία
 - β. αν αυξηθούν οι καταθέσεις κατά 1000 Ευρώ
 4. Αν είχατε μια αίτηση για δάνειο από έναν εργαζόμενο με σύμβαση ορισμένου χρόνου, ο οποίος εργάζεται για 40 χρόνια, που διαθέτει 5 πιστωτικές κάρτες, έχει καταθέσεις 20.000 Ευρώ, ζητάει δάνειο 15 έτη, είναι άγαμος και η συνολική βαθμολογία στο ερωτηματολόγιο που συμπλήρωσε ήταν 80 θα προτείνατε τη χορήγηση δανείου ή όχι; Ποια είναι η πιθανότητα χρεοκοπίας του συγκεκριμένου ατόμου;
 5. Η πιθανότητα χρεοκοπίας ενός συγκεκριμένου πελάτη ο οποίος άγαμος είναι 10%. Αν ένας έγγαμος πελάτης έχει ίδια όλα τα υπόλοιπα χαρακτηριστικά πόσο εκτιμάτε ότι θα είναι η πιθανότητα χρεοκοπίας του;
- Γ. Να κατασκευάσετε δένδρο ταξινόμησης με τη μέθοδο CRT χρησιμοποιώντας όλες τις διαθέσιμες μεταβλητές πρόβλεψης ποσοτικές και κατηγορικές. [Μονάδες 10]
 1. Ποιες μεταβλητές χρειάζονται στο προτεινόμενο μοντέλο για την πρόβλεψη της πιθανότητας να είναι εξυπηρετούμενο ένα δάνειο; Να δοθεί ο πίνακας ορθών ταξινομήσεων μέσω του προτεινόμενου δένδρου.
 2. Να απαντήσετε στο Ερώτημα B.4 με χρήση του προτεινόμενου δένδρου.
- Δ. Να διαμορφώσετε μοντέλο διαχωριστικής ανάλυσης χρησιμοποιώντας μόνο τις 5 ποσοτικές προβλέπουσες μεταβλητές των δεδομένων σας. Οι εκ των προτέρων πιθανότητες να υπολογισθούν με βάση τα ποσοστά καλών και κακών πελατών στα δεδομένα. [Μονάδες 10]

1. Να δοθούν οι τύποι των γραμμικών διαχωριστικών συναρτήσεων του Fisher και ο πίνακας ορθών ταξινομήσεων του
 2. Να απαντήσετε στο Ερώτημα Β.4 με χρήση του μοντέλου της διαχωριστικής ανάλυσης.
- Ε. Να συγκρίνετε τις τρεις διαφορετικές μεθόδους που αναπτύξατε στα Ερωτήματα Β, Γ, Δ αναφέροντας τα πλεονεκτήματα/μειονεκτήματα της καθεμίας. [Μονάδες 10]
- ΣΤ. Να εφαρμοσθεί η μέθοδος Backward προκειμένου να δημιουργήσετε ένα μοντέλο πρόβλεψης της ζημιάς που υφίσταται η τράπεζα από μη εξυπηρετούμενα δάνεια χρησιμοποιώντας όλες τις διαθέσιμες μεταβλητές, ποσοτικές και κατηγορικές. Να μην χρησιμοποιηθούν καθόλου όροι αλληλεπίδρασης μεταξύ των ανεξάρτητων μεταβλητών. Για την απάντηση στα επόμενα ερωτήματα μπορείτε να υποθέσετε ότι ισχύουν όλες οι απαραίτητες προϋποθέσεις για το κανονικό γραμμικό μοντέλο χωρίς να είναι απαραίτητο να ελεγχθούν. [Μονάδες 40]
1. Ποιες μεταβλητές είναι χρήσιμες για το μοντέλο πρόβλεψης; Να δοθεί ο τύπος υπολογισμού της ζημιάς και να σχολιασθεί η ποιότητα του μοντέλου.
 2. Να βρεθούν διαστήματα εμπιστοσύνης συντελεστή 95% για τις παραμέτρους του μοντέλου που αντιστοιχούν σε ποσοτικές μεταβλητές και να δοθεί η ερμηνεία τους.
 3. Ποια είναι η διαφορά στο ύψος της ζημιάς που προκαλείται από ένα μη εξυπηρετούμενο δάνειο που δόθηκε σε εποχικό εργαζόμενο από ένα μη εξυπηρετούμενο δάνειο που δόθηκε σε άτομο που εργάζεται με σύμβαση ορισμένου χρόνου με ίδια όλα τα υπόλοιπα χαρακτηριστικά; Να δώσετε άνω και κάτω όριο της διαφοράς για τα οποία είσατε βέβαιοι με πιθανότητα 95% .
 4. Ποια είναι η διαφορά στο ύψος της ζημιάς που προκαλείται από ένα μη εξυπηρετούμενο δάνειο που δόθηκε σε έγγαμο σε σχέση με άγαμο άτομο (τα δύο άτομα έχουν ίδια όλα τα υπόλοιπα χαρακτηριστικά); Να δώσετε άνω και κάτω όριο της διαφοράς για τα οποία είσατε βέβαιοι με πιθανότητα 95% .
 5. Αν είχατε μια αίτηση για δάνειο από έναν εργαζόμενο με σύμβαση ορισμένου χρόνου, ο οποίος εργάζεται για 40 χρόνια, που διαθέτει 5 πιστωτικές κάρτες, έχει καταθέσεις 20.000 Ευρώ, ζητάει δάνειο 15 έτη, είναι άγαμος και η συνολική βαθμολογία στο ερωτηματολόγιο που συμπλήρωσε ήταν 80 ποια θα είναι η ζημιά σε περίπτωση χρεοκοπίας; Ποια θα είναι η μέση ζημιά από άτομα με αυτά τα χαρακτηριστικά; Και στις δύο περιπτώσεις να δώσετε διαστήματα για τα οποία είσατε βέβαιοι με πιθανότητα 95%
 6. Για τις αιτήσεις για δάνειο από έναν εργαζόμενο με σύμβαση ορισμένου χρόνου, ο οποίος εργάζεται για 40 χρόνια, που διαθέτει 5 πιστωτικές κάρτες, έχει καταθέσεις 20.000 Ευρώ, ζητάει δάνειο 15 έτη, είναι άγαμος και η συνολική βαθμολογία στο ερωτηματολόγιο που συμπλήρωσε ήταν 80 ποια είναι η αναμενόμενη ζημιά; Για την πιθανότητα χρεοκοπίας να χρησιμοποιηθεί η τιμή που βρέθηκε στο Ερώτημα Β.4.
 7. Ένα στέλεχος της τράπεζας ισχυρίζεται ότι δεν υπάρχει διαφορά μεταξύ των απωλειών που προκαλούνται από ένα δανειολήπτη με σύμβαση ορισμένου χρόνου και ένα δανειολήπτη με σύμβαση αορίστου χρόνου με ίδια όλα τα άλλα χαρακτηριστικά, δίνοντας περιθώριο σφάλματος στον ισχυρισμό του 1%. Είναι σωστός ο συγκεκριμένος ισχυρισμός με βάση τα διαθέσιμα δεδομένα; Να αιτιολογήσετε την απάντησή σας.
 8. Να διαμορφώσετε ένα νέο μοντέλο το οποίο αποτελεί επέκταση του μοντέλου που ήδη χρησιμοποιήσατε στο Ερώτημα ΣΤ.1, έτσι ώστε να μπορέσετε να ελέγξετε κατά πόσο η διαφορά στο ύψος της ζημιάς που προκαλείται από ένα μη εξυπηρετούμενο δάνειο που δόθηκε σε έγγαμο άτομο σε σχέση με άγαμο (τα δύο άτομα έχουν ίδια όλα τα υπόλοιπα χαρακτηριστικά), επηρεάζεται από τις καταθέσεις ή από τον αριθμό των καρτών. Ποια είναι η απάντηση σε αυτό το ερώτημα με περιθώριο σφάλματος 5%;

Γενικές Οδηγίες

1. Η εργασία πρέπει να σταλεί ηλεκτρονικά με e-mail μέχρι την Κυριακή 12/7/2020.
2. Η συνολική έκταση της εργασίας δεν θα πρέπει να ξεπεράσει τις 10 σελίδες.