

פרויקט 3: סיכוני אשראי בבנק

תרגול נושא חוק בייס ומרחק מהלנוביס.

לקוח מתואר על ידי $A(x_1, x_2, \dots, x_6)$

x_1 – הכנסה

x_2 – חסכונות

x_3 – ערבות (בטחונות)

x_4 – גיל

x_5 – מקצוע (תנודתיות)

x_6 – היסטוריה פיננסית

סיכון נמוך: $P(C_1) = 0.8$

סיכון גבוה: $P(C_2) = 0.2$

תיאור קובצי נתונים

- א. קובץ של 100 לקוחות בסיכון נמוך
- ב. קובץ של 100 לקוחות בסיכון גבוה
- ג. קובץ של 100 לקוחות חדשים

מבנה הקבצים (solid, risky):

שורה 1: כותרות

שורה 2: משקלים (בהמשך הפיתוח תצטרפו אותם, בינתיים תקראו אותם ותאחסנו)

שורה 3 – 103: 100 לקוחות, לכל אחד 6 מאפיינים

תיאור התוכנה:

1. לימוד:

א. התוכנה תקלוט את קובץ א' ותחשב את סטטיסטיקת הלקוחות בסיכון נמוך C_1

(ממוצעים וסטיות תקן).

ב. התוכנה תקלוט את קובץ ב' ותחשב את סטטיסטיקת הלקוחות בסיכון גבוה C_2

(ממוצעים וסטיות תקן).

2. סיווג:

a. התוכנה תקלוט את קובץ הלקוחות החדשים ותסווג את הסיכון של כל לקוח חדש C_1 או C_2 . הסיווג יעשה על ידי חישוב 2 מרחקי מהלנוביס של כל לקוח חדש מסיכון

נמוך וגבוה לסיווג לסוג במרחק הנמוך יותר.

b. פלט:

i. סטטיסטיקת סיכון נמוך C_1 (ממוצעים וסטיות תקן).

ii. סטטיסטיקת סיכון גבוה C_2 (ממוצעים וסטיות תקן).

iii. סיווג הסיכון של 100 הלקוחות.

שאלות

1. מהו אחוז הלקוחות בסיכון גבוה שקבלת? האם זה הגיוני? נסה להסביר בעזרת משפט בייס וכן ערכי סטיות התקן השונות של הקבוצות השונות – בטוחים ומסוכנים (מה הקשר בין קבוצה עם סטיות תקן גבוהות ומידת המרחקים?).

2. בבנק אחר בדקו את הסטטיסטיקה של לקוחות בטוחים ולקוחות מסוכנים ומצאו את הנתונים הבאים:

לקוחות בטוחים	הכנסה	חסכונות	ערבות (בטחונות)	גיל	היסטוריה פיננסית	קביעות מקצועית
ממוצע	30000	150,000	250000	50	10	1
סטית תקן	7000	40000	70000	12	2.5	0.25

לקוחות מסוכנים	הכנסה	חסכונות	ערבות (בטחונות)	גיל	היסטוריה פיננסית	קביעות מקצועית
ממוצע	5000	50000	100000	30	6	0.7
סטית תקן	1000	10000	10000	5	0.5	0.1

לקוח צעיר הגיע לבנק עם הנתונים הבאים:

הכנסה	חסכונות	ערבות (בטחונות)	גיל	היסטוריה פיננסית	קביעות מקצועית
2500.00	25000.00	50000.00	15.00	3.00	0.35

א. חשב את מרחק נתוני הלקוח מסטטיסטיקת לקוחות בטוחים וכן מסטטיסטיקת לקוחות מסוכנים.

ב. סווג את הלקוח (בטוח או מסוכן?)

ג. האם סיווג הלקוח הגיוני עבור הנתונים שלו? הסבר.

הערה: הרץ את הנתונים בקוד שכתבת ומצא המרחקים.

3. נסה להציע שיטה למדידת רמת סיכון של לקוח והחלטה על גובה הריבית עבורו. כלומר, במקום לסווג לקוח חדש ל "בטוח" או "מסוכן", נסה להעריך את מידת הסיכון שלו (ולכן את הריבית הנכונה עבורו, שמפצה על הסיכון שלו).