

נושא הפרויקט - חיזוי נטישת לקוחות בנק

בפרויקט זה נחקר את מאגר הנתונים של לקוחות בנקים, המכיל מידע פיננסי ודמוגרפי (כגון דירוג אשראי, מין, גיל, ותק, משכורת וכו') עבור אלפי לקוחות. המטרה היא לחזות האם לקוח ינטוש את שירותי הבנק על פי מספר פרמטרים תוך שימוש בכלי למידת מכונה שונים שלמדנו בכיתה.

מאגר הנתונים - <https://www.kaggle.com/datasets/anandshaw2001/customer-churn-dataset>

שאלות

- מהם המשתנים המשפיעים ביותר על נטישת לקוחות בבנק (כגון גיל, יתרה בחשבון, משך התקשרות, הכנסה חודשית וכו')?
- איזה אלגוריתם מספק את הדיוק הגבוה ביותר בזיהוי לקוחות שעומדים לנטוש את שירותי הבנק?
- כיצד פעולות הנדסת תכונות (כגון הורדת מימד, נרמול, יצירת מאפיינים חדשים) משפיעות על דיוק המודלים?

מודלים

- Decision Trees - יישום מודל עץ החלטה כדי להבין מהם הגורמים המשמעותיים ביותר לנטישת לקוחות.
- SVM - בדיקת יכולת ההפרדה של מודל SVM לינארי ולא לינארי בין לקוחות שנוטים ללקוחות שנשארים.
- Adaboost: שימוש באלגוריתם הגברה כדי לשפר את ביצועי המודלים הפשוטים ולהפחית טעויות סיווג.
- k-NN - בדיקת ביצועי k-NN והשוואתו לאלגוריתמים אחרים על סמך דמיון בין לקוחות שונים.

תהליך העבודה יכלול:

- שלב מקדים - ניקוי הנתונים, טיפול בערכים חסרים, נרמול והמרת משתנים קטגוריאליים.
- ניתוח ראשוני (EDA) - חקירת התפלגויות המשתנים, חישוב סטטיסטיקות והצגת ויזואליזציות להבנת קשרים פוטנציאליים.
- יישום המודלים - יישום והשוואת ארבעת האלגוריתמים על בסיס מדדי ביצועים כמו דיוק, F1-Score, Precision, Recall, Accuracy.
- הסקת מסקנות - זיהוי המודל האופטימלי על סמך איזון בין דיוק, פשטות וזמן ריצה, וגיבוש המלצות לבנק להתמודדות עם נטישת לקוחות.

מיכאל מורלי 318410941

נטלי סדיקוב 206396863