תרגיל 4 – with Python Machine Learning

בתרגיל זה תתנסו ביישום אלגוריתמי למידת מכונה תוך שימוש בשפת Python לטובת משימת סיווג.

עליכם להשתתף בתחרות Loan Prediction שבאתר Analytics Vidhya.

<https://datahack.analyticsvidhya.com/contest/practice-problem-loan-prediction-iii/>

חברת Dream Housing Finance מעוניינת למכן את תהליך אישור בקשות ההלוואה. עליכם לנבא אם בקשה תאושר על סמך יתר הפרטים שסופקו בטופס הבקשה המקוון.

השתמשו בסט האימון שבאתר, המכיל פרטים שונים אודות מבקשי ההלוואות, וסווגו כל בקשה כמאושרת (כן/לא). לרשותכם תרגיל מודרך (HW4.ipynb) שיסייע לכם בפתרון השלבים הראשונים של התרגיל. התחילו בפתרון התרגיל המודרך ומשם התקדמו ונסו ללמוד מודלים חכמים יותר בכוחות עצמכם.

מבנה הציון:

* 5% מיקומכם היחסי בטבלת התוצאות.
* 95% שלמות ותקינות התהליך המתואר בדוח המסכם.

**מבנה הדוח המסכם:**

* תעדו את התהליך והקוד. השתמשו לשם כך בJupyter Notebook (קובץ ipynb), הממוקם בgithub שלכם.
* עליכם להגיש לאתר Analytics Vidhya לפחות 2 ניסיונות סיווג, תוך שימוש במסווגים שלא הופיעו בתרגיל המודרך.
* ציינו מהו מיקומכם בטבלת הleaderboard, ציינו שם משתמש באתר Analytics Vidhya וצרפו צילום מסך של לוח התוצאות המראה את מיקומכם בעת ההגשה.
* לפני כל הרצה כווננו פרמטרים חשובים
* על הדוח המסכם לכלול עבור כל הגשה שבוצעה לאתר Analytics Vidhya:
  + תיאור העיבוד המקדים (במידה ובוצע)
  + תיאור האלגוריתם בו השתמשתם
  + תיאור תהליך כוונון פרמטרים חשובים (כולל תיאור גרפי)
  + הפניה לקובץ שהוגש (שישב גם הוא בgithub)
  + צילום מסך של הדירוג שניתן לכם עבור ההגשה (צלמו את המסך שהתקבל לאחר ההגשה לאתר analyticsvidhya)

**דגשים:**

* כל זוג יצור חשבון משתמש ב Analytics Vidhya כדי להוריד את הנתונים הרלוונטיים ולהשתתף בתחרות. ההסברים המלאים נמצאים באתר.
* תוכלו להיעזר ברעיונות מתוך פורומים המלווים את התחרות, אך יודגש כי אין להעתיק פתרונות מהאינטרנט.

יש להגיש את התרגיל בזוגות עד 14/1/2018.

תרגילים שיוגשו באיחור "יקנסו" באופן הבא: 5 נקודות יורדו מציון התרגיל לכל יום איחור.

ההגשה תתבצע בתיקיה המתאימה באתר הקורס באופן הבא:

1. על כל אחד מבני הזוג להגיש את העבודה.

2. יש להגיש קובץ ובו קישור למאגר ה Githubשל התרגיל וכן שמות השותפים לתרגיל. שם הקובץ יהיה מהסוג ID1\_ID2.txt (תעודות הזהות של המגישים מופרדות באמצעות קו תחתון).

בהצלחה ובהנאה!