**כריית מידע – מטלה 2**

# Implement the pre-process

במטלה זו, ניקח את המידע שקיבלנו בחלק א' וסידרנו בעזרת כל השיטות שלמדנו בחלק הראשון של הקורס, ונריץ עליהם מספר מודלים של מסווגים שונים שיתאמנו עליהם, נחפש את אלו שלדעתינו וביחס למשימה מקבלים את התוצאות הטובות ביותר ביחס לTEST ובסוף את המסווג הטוב ביותר נריץ על המידע שקיבלנו הפעם קובץ הTEST שלא ראינו קודם, ונקווה לחזות בצורה הטובה ביותר.

**הכנת המידע:**

* למרות שזה חלק ששמור לסוף, מכיוון שידענו שקובץ הTEST גם צריך להיות מסודר כראוי, ניגשנו אליו בראשונה לעשות תהליך pre-proccess , כמובן שעבדנו לפי אותם שיטות שעשינו להכנת המידע על חלק א'.   
  (את העבודה לחלק זה עשינו בקובץ **clean\_test\_data.ipynb**)
* לאחר מכן, חזרנו לעשות שיפוץ קטן ומתומצת על המידע בחלק א' שלנו, עם אותם שיטות בדיוק, פשוט וידאנו שביצענו פעולות דומות עבור השלמות המידע והנרמול כדי לדייק ככל הניתן בתהליך הסיווג והחיזוי פה במטלה זו.  
  (את העבודה לחלק זה עשינו בקובץ **clean\_train\_ex2.ipynb**)

את הקבצים ייצאנו לשני CSV מוכנים לעבודה בשם:

* **test\_ready.csv** עבוד המידע TEST המוכן
* **test\_ready.csv** עבוד המידע TRAIN המוכן

ומכאן ואילך למען הסדר הטוב, את שאר המטלה עבדנו על קובץ בשם **Ex\_2\_full.ipynb**

ראשית, ייבאנו והתקנו כמובן את כל הספריות הנדרשות למטלה זו.

לאחר מכן, ייבאנו את הקובץ המוכן לעבודה של הTRAIN

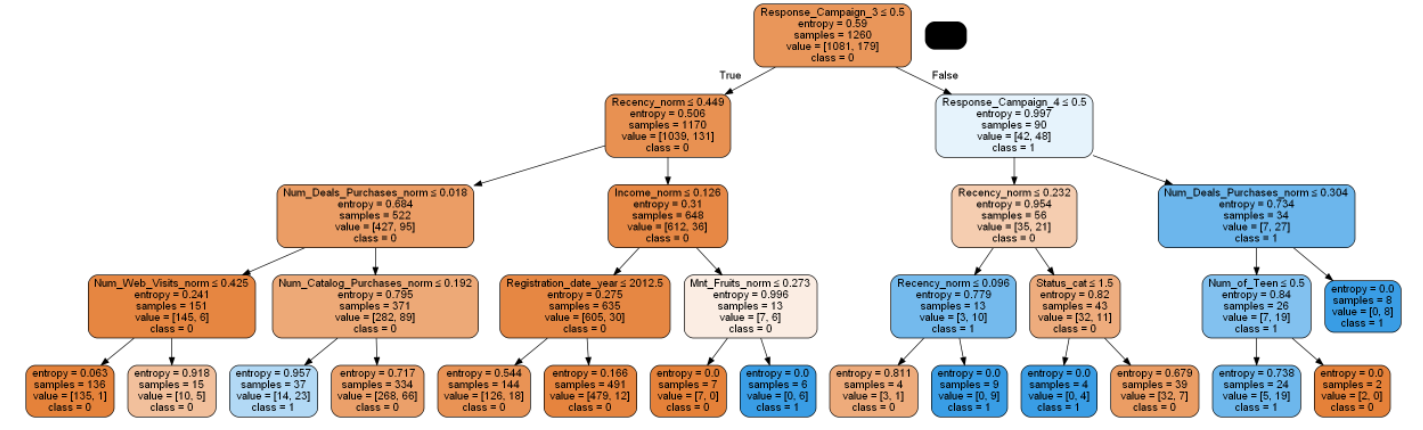
הגדרנו את המידע ואת ערך המטרה בנפרד.

וכן פיצלנו את כל הDATA שלנו לשני חלקים , train-test ביחסים של 75-25 אחוז

**תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי**

תמונה של עץ החלטה שיצא לנו מהמסווג הראשון:



ויזואליזציה בהקשר של יער של עצי החלטות:

