# מבחן

אנו מעוניינים למצוא את האנשים החשודים כמחבלים מתוך מאגר של אנשים שיש לנו. נעשה זאת על ידי בנייה של מודל Machine Learning המשתמש במידע אודות האנשים במאגר שזמין לנו. המשימה שלכם היא לבנות (חלק מ) **תשתית הפייפליין** עבור הפרויקט. להלן תיאור הפייפליין והדרישות מהתשתית, כפי שהחוקרים רואים את זה.

## הנחיות

1. מותר להשתמש באינטרנט
2. יש לעשות את המבחן לבד
3. יש לכתוב את הקוד בפייתון ולהגיש את כל מה שנדרש בשביל שהפרויקט ירוץ

## דרישות שהתקבלו מהחוקרים

הפייפליין בנוי משני שלבים עיקריים. ראשית, מחשבים פיצ'רים על כל אחד מהאנשים. הoutput של שלב זה הוא pandas dataframe שכל שורה בו מייצגת בן אדם מתוך המאגר, וכל עמודה היא פיצ'ר שחושב על אותו אדם. בשלב השני של הפייפליין, מודל כלשהו מקבל את הdataframe ורץ עליו. המידע נלקח מתוך קבצי csv, בהם יש מידע על האנשים שיש במאגר ועל המשפחות שלהם. דוגמאות לפיצ'רים:

* גיל
* לכמה חשודים התקשר
* כמות השיחות מעל x דקות (x = 2, 4, 6)
* כמה שיחות קיבל

התשתית נדרשת להיות קונפיגורבילית בהיבטים הבאים:

* איזה פיצ'רים רצים
* על איזה אנשים הפיצ'רים מחושבים
* פרמטרים (כולל בתוכו היפרפרמטרים של המודל, וכן פרמטרים שמועברים לפיצ'רים)

## הגדרת המשימה

אתם נדרשים לממש את התשתית עבור השלב הראשון של הפייפליין (חישוב הפיצ'רים).

1. עבדו עם הdata שיש בקובץ הzip המצורף
2. הגדירו את התממשקות הפיצ'רים לתשתית (הפיצ'רים נכתבים על ידי החוקרים)
3. ממשו את הפיצ'רים לדוגמא באופן שהגדרתם
4. כתבו את קובץ הmain של הפייפליין, שמקבל את קונפיגורציית ההרצה (אתם מגדירים איך הקונפיגורצייה מוגדרת ומועברת)
5. כתבו את תוצאת הפייפליין (הdataframe של תוצאות הפיצ'רים) למקום שנראה לכם מתאים.
6. הפייפליין הזה אמור לרוץ פעם בשעה ולכתוב את התוצאות לDB רלציוני. הסבירו מה נדרש לעשות על מנת להעלות את הפייפליין הזה לproduction.

## דגשים:

* המטרה כאן היא לעצב את התשתית באופן שיהיה נוח לשימוש ויאפשר גמישות רבה. התשתית צריכה להיות גמישה בפיצ'רים שהיא מריצה, האנשים מתוך המאגר שהיא רצה עליהם, וההיפרפרמטרים שמשתמשים בהם. כמובן שלא חייבים להסתפק באלה וניתן לחשוב על רעיונות נוספים.
* אם אתם מכירים טכנולוגיות שיכולות לעזור לכם - מעולה, אתם מוזמנים להשתמש. אם אתם מכירים אבל בכל זאת מעדיפים לממש הכל בעצמכם (בגלל overhead, משך זמן התרגיל…) הסבירו במה ניתן להשתמש ומה השיקולים שהיו מנחים אתכם בבחירת טכנולוגיה מסויימת.
* קחו בחשבון שהפיצ'רים שהחוקרים ישתמשו בהם יהיו יותר מסובכים וכן יהיו כבדים יותר מבחינה חישובית.
* חשוב שלחוקר יהיה נוח להשתמש בתשתית ושהוא יבין מה קורה בכל שלב של הריצה.