



שם הפרויקט		מס' פרויקט
משימת חיזוי – פרק הזמן שלוקח לעבור מתזרים שלילי לחיובי		2022-01-208
מנחה שותף	מנחה אקדמי	
פרופ' גלעד רביד	ד"ר נמרוד טלמון	
חברי הצוות		
	יובל שמר	ענת קוטליאר
	shemeryu@post.bgu.ac.il	anatkot@post.bgu.ac.il

תקציר

נושא ומטרה: הפרויקט עוסק במשימת למידה מסוג חיזוי עבור שירות התזרים של חברת RiseUp. חברת RiseUp הינה חברת סטארט אפ בתחום ה-fintech, המייצרת פלטפורמה בה היא מציגה ללקוחותיה את פירוט התנועות הכספיות שלהם בפילוח לתחומים (מזון, רכב, הוצאות קבועות ועוד), במטרה ליצור ללקוח שקיפות מלאה לגבי התנהגותו הכלכלית ובכך לעזור להם לפתח מסוגלות עצמית ולצמוח כלכלית. השאלה המרכזית הנבחנת בפרויקט בעזרת משימת הלמידה היא - האם ההצטרפות לשירות RiseUp "משפרת" את התזרים החודשי ומעלה את ההסתברות "לנצח את החודש". כלומר, האם הפער בין ההכנסות לבין ההוצאות (קבועות ומשתנות) הופך מחיובי לשלילי, מהו פרק הזמן הנדרש לכך, ואילו מרכיבים השפיעו על השינוי.

שיטה: בחרנו לבנות מודל חיזוי החוזה את מספר החודשים הנדרש ללקוח להגיע מתזרים שלילי לחיובי (יותר הכנסות מהוצאות) מרגע קבלת התזרים הראשוני. פיתוח משימת הלמידה בוצע בשפת התכנות python, תוך איסוף הנתונים הגולמיים מבסיס הנתונים של החברה, בחינת הקורלציות אל משתנה המטרה, חילוץ פיצ'רים חדשים ובחירת בעזרת כלים סטטיסטיים, ובשלב האחרון - אימון המודל, תוך בחינת שלושה מודלי למידה: Decision tree, Artificial neural network, SVM.

כאשר לקוח חדש נרשם לשירות RiseUp המערכת מקבלת את תנועות החשבונות שלו חודשיים אחורה ובונה לו תזרים ראשוני, בכל חודש בו נמצא הלקוח בשירות נוספות תנועות החשבון שלו ועל פיהן המערכת בונה תזרים עדכני המתחשב בכלל חודשי העבר. מתוך ההבנה כי עם הצטרפות לקוח חדש אל השירות למערכת אין מידע רב אודותיו והיא אוספת מידע נוסף בכל חודש וכך מגיעה לדיוק גבוה יותר בחרנו גם אנחנו לבנות שלוש מכונות למידה אשר חוזות את מספר החודשים שיארך ללקוח להגיע לתזרים חיובי, כל מכונת למידה מבצעת את החיזוי בנקודת זמן אחרת בעזרת כמות נתונים שונה. השערתנו היא שככל שלמכונת הלמידה יותר דאטה כך אחוזי הדיוק יגדלו.

תוצרים: לאחר בחינת שלושת מודלי הלמידה תוך ביצוע Hyperparameter tuning אחוזי הדיוק הטובים ביותר התקבלו עבור מודל ה-SVM, עבור שלושת המכונות. בנוסף נוכחנו לגלות כי אחוזי הדיוק הנמוכים ביותר התקבלו במכונה 0 ועלו בהתאמה, כך שלמכונה 2 57.39% דיוק, דבר המאשש את השערתנו כיוון למכונה זו כמות דאטה הגדולה ביותר.

מילות מפתח: תזרים, מודל, חיזוי, דאטה, אימון מודל.