

# *IML* האקתון

9 ביוני 2023

Mission 1 - Agoda

לוגינים של המגישים:

michaelmerzin,323913277

yonmiroshnik,318209939

ofir.rosen,325538189

sean.cherny,213716608

פרויקט זה חולק לכמה חלקים.

### 1.2.1 תחזית ביטול הזמנה

בחלק זה קיבלנו בעצם סט אימון  $S$ , פירקנו אותו ל- $X, y$ , כאשר ה- $X$  היא מטריצת נתונים על הזמנות למלונות, וה- $y$  היו במקור תאריכי ביטול הזמנה אם הייתה כזו. נזכור כי כאן ערכי ה-Label-ים הם בדידים:  $\mathcal{Y} = \{0, 1\}$ . העברנו עליו פרי-פרוסס בו עשינו את הדברים הבאים:

- ראשית, פיצלנו את הנתונים באמת למטריצה  $X_{\text{train}}$  ווקטור  $y_{\text{train}}$ .
- כעת, את העמודה  $y_{\text{train}}$  במקום שהיא תכיל תאריכי ביטול עשינו שהיא תכיל משתנה בינארי של האם היה ביטול, שהרי אנחנו רוצים קלאסיפייר שבדק האם היה ביטול בעלמא. החלטנו ש-0 זה שאין ביטול ו-1 זה שהיה ביטול.
- עמודות של תאריכים כמו  $checkin\_date$  או  $booking\_date$  החלפנו בעמודות של יום בשנה ובשנה.
- הוספנו עמודה  $stay\_duration$  שמתארת את משך הנופש שהוזמן. נתון על פי הנוסחה:  
$$stay\_duration = checkout\_date - checkin\_date$$
כאשר זהו חיסור תאריכים והמימוש של מימוש תאריכים לא כזה טריוויאלי אבל עשינו את זה.
- את העמודה  $checkout\_date$  מחקנו כי היא כשלעצמה פחות חשובה.
- לעמודות קטגוריאליות כגון  $charge\_option$  או  $language$  עשינו  $get\_dummies$ .
- בדקנו תקינות של ערכים כפי שהוגדר בתרגיל. בקבוצת האימון - מחקנו שורות עם ערכים לא תקינים. בקבוצות בדיקה לא מחקנו שורות, כי אסור, אלא נתנו ערכים דיפולטיביים (כמו 0 או הערך השכיח וכו').
- מחקנו את העמודות הבאות, כיוון שלאחר בדיקות הן נראו לא רלוונטיות לסיכויי הביטול:  
 $hotel\_id, hotel\_live\_date, hotel\_brand\_code, h\_booking\_id$

לאחר הפרי-פרוסס שלחנו את הנתונים ל-Adaboost. השתמשנו במודל זה כי הוא קלאסיפייר טוב. המודל נתן לנו תחזיות קלאסיפיקציה על סט הבדיקה והבאנו אותן בקובץ נפרד.

### 1.2.2 תחזית הפסדים מביטול הזמנה

בחלק זה עשינו פרי-פרוסס דומה. אבל הפעם נתנו את הנתונים למודל רגרסייה לינארית כי כאן ערכי ה-Label-ים צריכים להיות רציפים  $\mathcal{Y} = \mathbb{R}$ . המודל נתן לנו תחזיות למחירים על סט הבדיקה והבאנו אותן בקובץ נפרד.

### 1.2.3 חלק זה מפורט ב- $pdf$ נפרד.

### 1.2.4 חלק זה מפורט ב- $pdf$ נפרד.