**אימון מודל נוסף לבחירתכם**

בחרו מודל סיווג נוסף מכל סוג שתרצו )חפשו באופן עצמאי(, אמנו וכווננו את הפרמטרים שלו.

פרטו בקצרה על מנגנון הסיווג שלו, על הפרמטרים שבחנתם ובחרתם, והציגו את ביצועיו

בתור מודל נוסף בחרנו להשתמש במודל Logistic Regression

**מנגנון הסיווג של Logistic Regression**

Logistic Regression הוא מודל סיווג שבבסיסו מודל לינארי. הוא משתמש בפונקציה לוגיסטית על מנת להחזיר ערך הסתברות עבור כל מחלקה, כאשר ההחלטה הסופית מתקבלת לפי ערך הסף (threshold) שנקבע מראש (בדרך כלל 0.5). הפונקציה הלוגיסטית מתוארת על ידי הנוסחה הבאה:

z הוא צירוף לינארי של המשתנים המסבירים (features):

**הפרמטרים שבחנו ובחרנו**

בעת אימון Logistic Regression, יש מספר פרמטרים חשובים שיש לכוון על מנת לשפר את ביצועי המודל:

פרמטר **C** קובע את עוצמת הרגולריזציה של המודל. רגולריזציה היא תהליך שמסייע למנוע overfitting על ידי הוספת עונש למשקלים גדולים במודל.

* **ערכים קטנים של (0.0001(** מתאימים לעונש רגולריזציה חזק יותר, כלומר מניעת overfitting.
* **ערכים גדולים של (10000(** מתאימים לעונש רגולריזציה חלש יותר, כלומר הדגש על התאמה מדויקת של המודל לנתונים הקיימים.

פרמטר**penalty** קובע את סוג הרגולריזציה שתתבצע:

* **'l1'** : רגולריזציה המבוססת על הערך המוחלט של המשקלים, מה שמוביל לעונש על מספר קטן של משקלים גדולים, ויכולה לגרום לחלק מהמשקלים להיות אפס.
* : **'l2'**רגולריזציה המבוססת על ריבוע המשקלים, מה שמוביל לעונש על מספר רב של משקלים קטנים.
* **'elasticnet'** : שילוב של רגולריזציה 'l1' ו-'l2'.
* **'none'** : ללא רגולריזציה.

פרמטר **solver** קובע את האלגוריתם שמשמש לאופטימיזציה. לכל אלגוריתם יש יתרונות וחסרונות, בהתאם לבעיה ולמחשב שבו משתמשים:

* **'newton-cg'** : מתאים לבעיות גדולות ומשתמש ב-gradient descent עם הסקה ניוטונית.
* **'lbfgs'** : גרסה של gradient descent שנקראת BFGS, מתאימה לרוב הבעיות.
* **'liblinear'** : אלגוריתם מבוסס על coordinate descent, מתאים לבעיות קטנות ובינוניות.
* **'sag'** : (Stochastic Average Gradient ) מתאים לבעיות גדולות.
* **'saga'** : שיפור של sag, מתאים גם לרגולריזציה 'elasticnet'.

פרמטר **max\_iter** קובע את מספר האיטרציות המקסימלי שהאלגוריתם ירוץ בו עד לעצירה. ערכים גבוהים יותר מאפשרים לאלגוריתם לרוץ זמן רב יותר ולשפר את ההתכנסות, אך גם עלולים להאריך את זמן החישוב.

**ביצועי המודל**

לאחר כיוונון הפרמטרים, ניתן להעריך את ביצועי המודל באמצעות מספר מדדים, כמו:

F1 Score: הממוצע ההרמוני של Precision ו-Recall.

מודל Logistic Regression הוא כלי סיווג חזק ופשוט לשימוש, במיוחד כשמכוונים נכון את הפרמטרים שלו. בחירת הפרמטרים המתאימה והערכת הביצועים באמצעות מדדים סטטיסטיים מאפשרים לנו לקבל תובנות מעמיקות יותר על המודל ועל ביצועיו. .

פונקציית feature\_importances\_ במודל DecisionTreeClassifier מחזירה מדדים של חשיבות המאפיינים במודל. כל ערך במערך זה מייצג את המשקל או החשיבות של כל מאפיין בהגדרת המודל שנוצרה על ידי עץ ההחלטה.

החשיבות של מאפיין מציינת כמה חשוב המאפיין הזה היה בתהליך הלמידה של המודל. ככל שהערך של החשיבות גבוה יותר, כך המאפיין הוא יותר חשוב לקביעת החלטות במודל.

התוצאות של פונקציית feature\_importances\_ יכולות להיות שימושיות להבנת כיצד המודל קיבל את ההחלטות שלו. בהקשר של הסעיף הקודם, שבו חקרנו את שימוש באלגוריתמים שונים לבעיות אופטימיזציה מרובות סוגיות ובדקנו תוצאות בהתאם למספר הסוגיות והסוגיות השונות.

feature importances

3 Type of Travel 0.223052

7 Inflight wifi service 0.214225

18 Cleanliness 0.162774

1 Customer Type 0.121908

17 Inflight service 0.076073

9 Ease of Online booking 0.033348

13 On-board service 0.024158

15 Baggage handling 0.023126

12 Seat comfort 0.022885

16 Checkin service 0.017747

5 Flight Distance 0.016846

14 Leg room service 0.015798

2 Age 0.013996

10 Gate location 0.007576

4 Class 0.006399

20 Arrival Delay in Minutes 0.005046

11 Food and drink 0.004007

6 Plane colors 0.003950

8 Departure/Arrival time convenient 0.003426

19 Departure Delay in Minutes 0.002452

0 Gender 0.001211

לדוגמה, ניתן לראות ש-"Type of Travel" ו-"Inflight wifi service" מופיעים בקרב המאפיינים הכי חשובים, עם ערכי חשיבות גבוהים מאוד של כ- 0.22 ו- 0.21 בהתאמה. מאפיינים אלו יכולים להיות קריטיים להגדרת התגובה של המודל בהתאם לסוג הנסיעה ולשירותי האינטרנט במהלך הטיסה.

בהשוואה למסקנות מהסעיף הקודם שבו בדקנו אלגוריתמים לבעיות אופטימיזציה מרובות סוגיות, ניתן להגיד ששימוש במאפיינים כמו סוג הנסיעה ושירותי האינטרנט במהלך הטיסה משקפים את הגישה להכרת תבניות ולקבלת החלטות במודל המותאם לפי פרטי הבעיות.