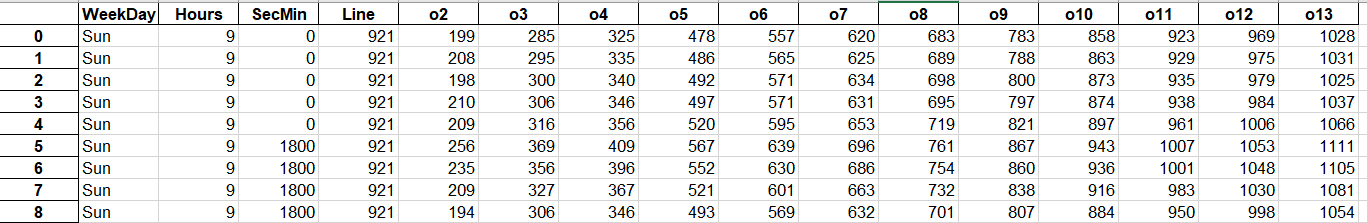
מסמך הסבר:

אתן הסר לקופץ פייטון **XGBOOST\_Regression.ipynb.** זה קובץ של מודל שהיא רגרסיה XGBOOST. הרעיון המרכזי של מודל היא שכל תחנה היא מודל שכל פעם שאנו מתקדמים לתחנה הבא אנו מגדלים כמות הכניסות במלים אחרות אנו מסתקלים על דאטה של תחנה קודמת, בעצם כל תחנה אחרי תחנה ראשונה תקבל X של תחנות קודמות. בסופו של דבר הרעיון היה להגיע לזמנים יותר מדויקים ולצאת מנקודת הנחה שבתחנות ראשונות דיוק הוא הכי טוב כי זה קרוב לתחנה מרכזית וכל הבעיה היא לנבעה משהו שהוא במרחק הרבה יותר רחוק מתחנה מרקזית מתחנה לדוגמה תחנה 6. בסופו של דבר המודל הזאת מנבעה את כל המסלול מתחנה ראשונה עד תחנה אחרונה שאנו חוקרים אותה. הדיוקים אפשר לראות בטבלה סופית:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Model Statistic data information with weekday | | | | | |  |  |  |  |
|  | **Station 2** | **Station 3** | **Station 4** | **Station 5** | **Station 6** | **Station 7** | **Station 8** | **Station 9** | **Station 10** | **Station 11** | **Station 12** | **Station 13** |
| R^2 | 0.39 | 0.73 | 0.89 | 0.9 | 0.9 | 0.94 | 0.97 | 0.98 | 0.98 | 0.98 | 0.99 | 0.97 |
| RMSE | 38.2 | 33.6 | 23 | 24 | 31 | 25 | 16 | 14 | 14 | 16 | 10 | 20 |
| The statistic Data: difference between y-test to predicted | | | | | | | | | | | | |
| Min | -202 | -116 | -112 | -69 | -91 | -169 | -68 | -71 | -139 | -217 | -44 | -223 |
| Max | 196 | 166 | 136 | 90 | 184 | 108 | 136 | 49 | 60 | 32 | 58 | 67 |
| STD | 23 | 22 | 12 | 15 | 18 | 15 | 9 | 9 | 8 | 7 | 7 | 10 |
| Average | 38.2 | 33.6 | 23 | 24 | 31 | 25 | 16 | 14 | 14 | 16 | 10 | 20 |
| Hypothesis Accuracy for Tuesday 11:25 | | | | |  |  |  | Hypothesis Accuracy for Tuesday 16:45 | | | | |
| H0 | 81% | |  |  |  |  |  | H0 | 81% | |  |  |
| H1 | 19% | |  |  |  |  |  | H1 | 19% | |  |  |
| Hypothesis Accuracy for whole days | | | | |  |  |  | Hypothesis Accuracy for whole days | | | | |
| H0 | 91% | |  |  |  |  |  | H0 | 79% | |  |  |
| H1 | 9% | |  |  |  |  |  | H1 | 21% | |  |  |
|  |  |  |  |  | Predicted Data | | |  |  |  |  |  |
| 11:25 | 190 | 265 | 307 | 449 | 528 | 586 | 652 | 752 | 826 | 893 | 943 | 987 |
| 16:45 | 256 | 416 | 451 | 628 | 758 | 866 | 938 | 1031 | 1107 | 1167 | 1198 | 1269 |
|  |  |  |  |  | Actual Data H0 | | |  |  |  |  |  |
| 11:25 | 182 | 252 | 292 | 437 | 513 | 574 | 638 | 745 | 822 | 886 | 931 | 975 |
| 16:45 | 243 | 387 | 428 | 613 | 757 | 834 | 899 | 999 | 1072 | 1137 | 1184 | 1232 |
|  |  |  |  |  | Actual Data H1 | | |  |  |  |  |  |
| 11:25 | 177 | 248 | 298 | 353 | 400 | 446 | 489 | 572 | 634 | 676 | 701 | 752 |
| 16:45 | 247 | 333 | 366 | 587 | 650 | 746 | 854 | 941 | 996 | 1037 | 1092 | 1113 |

כמו שאפשר לראות זה נותן דיוקים ממש טובים. יש פה גם בדיקת השארות האם יש הבדל בין זמן מנובה לזמן אמיתי וראינו גם פה דיוקים ממש טובים. אפילו שאנחנו מקבלים את H1 גם פה תוצאות לא רעות אבל אנחנו שואפים לתוצאה מושלמת. אני חושב שהצלחתי להעביר תמונה כללית של מודל, בגדול אפשר היה לעשות מודל הרבה יותר חכמה, הרעיון הזה הגיע יותר מאוחר לקראת סוף פרויקט. אפשר היה לשלות בעומק של עץ ולכל תחנה כל פעם לבחור מודל שונה ולא להשתמש באותה מודל כל הזמן. מה הכוונה פה: לדוגמה אני אקח תחנה שהיא מספר 6 ודיוקים הם בהתם למודל שיש לי עכשיו, היתי מנסה לשפר את התוצאות בעזרת ערכים אחרים עבור מודל עד שהיתי מגיע למשהו יותר טוב, היתי חייב להגביל את הבדיקות כדי לא להגיע לOUTLAYERS.

בקוד באפשר לראות התנגות של כל מודל תחנה באזרת גרף. כל התוצאות והסברים עבור פונקציות וקטעי קוד מופעים בקופץ פייטון. בקוד אפשר לשחק אם זמנים. בכניראשית שהיא משותפת לכל המודלים אנו תמיד שמים זמן יציאה מתחנה מרכזית, גם מודל תלויה יום בשבוע מיום ראשון עד יום חמישי לפי הסדר. כל מה שצריך להיות בקוד הזה מצורף בתקיה, בוודאי מצורף קופץ סופי של כל הדאטה סופי שאספנו, בעצם בקובץ:



אז אפשר לראות יום בשבוע, זמן יציאה שעות ודקות שהפחמו לשניות(זה לא שינה לנו דיוק של מודל), מספר קו, כל הנתונים האלו הם משותפים לכל המודלים מתחנה 2 עד תחנה 13. אחרי זה יש עמודות עבור כל תחנה והזמנים שלה, כמה זמן לוקח להגיע לתחנה לפי עמודה, ויש כאלו 12 עמודות. בסופו של דבר כל עמודה תהפוך לפיצירים עבור תחנות הבאות לדוגמה תחנה o4 בנוסף ל X משותף גם תקבל כ X את o2,o3.

כל הקודים נבדרו והם עובדים, לכל קטע קוד יש הסבר ואופן עבודה שלו, אפשר לשחק ולבדוק זמנים אחרים עבור מודל שבניתי.