

Percenatges of NaN Values For Every Attribute resultdir Missing Ratio % resultspeed preciptotal avgspeed dewpoint tmax stnpressure attributes tmin wetbulb cool heat tavg sealevel snowfall sunset sunrise codesum depart 30 40 50 10 20 NaN Values percentages

השלמת מידע חסר

- ♦ השלמת מידע חסר הוא חלקקריטי בעיבוד המידע
- ♦ הוויזואליזציה הבאה עזרה לדעתבאילו תכונות להתמקד

Heatmap of weather Data Numeric Features -0.32 -0.12 -0.2 -0.13 0.39 -0.93 0.76 -0.21 0.013 0.06 -0.9 0.81 0.98 -0.45 0.0063 0.98 -0.19 0.055 0.93 0.98 -0.94 0.8 -0.44 0.72 0.1 -0.34 -0.086 -0.23 -0.097 0.98 -0.88 0.73 -0.42 -0.15 0.15 -0.93 0.76 0.69 -0.18 0.11 0.17 -0.34 -0.45 -0.94 -0.88 -0.53 -0.65 0.23 -0.069 -0.087 0.33 -0.93 0.73 0.76 -0.33 0.61 -0.071 0.013 0.092 -0.27 -0.027 -0.16 -0.048 0.26 -0.69 0.084 -0.071 0.14 0.24 0.024 0.061 0.038 0.038 -0.33 -0.14 0.067 -0.038 -0.25 -0.04 -0.19 -0.0061 -0.024 1 0.096 -0.031 -0.02 0.049 -0.0064 0.055 -0.063 -0.15 -0.18 0.23 -0.071 0.084 -0.14 0.11 -0.069 0.013 -0.071 0.067 0.096 1 0.022 -0.17 -0.0017 -0.063 0.054 -0.026 0.24 0.17 -0.087 0.092 0.14 -0.038 -0.031 0.022 1 0.2 -0.011 -0.033 -0.14 -0.19 sealevel -0.0063 -0.32 -0.33 -0.34 -0.31 -0.34 0.33 -0.27 0.24 -0.25 -0.02 -0.17 0.2 -0.19 -0.078 -0.27 -0.41 0.1 -0.027 0.024 -0.04 0.049 -0.0017 -0.011 -0.19 0.14 -0.022 0.22 -0.16 0.061 -0.19 -0.0064 -0.063 -0.033 -0.078 0.12 -0.23 -0.23 -0.26 -0.25 avgspeed -0.0032 -0.13 -0.07 -0.097 -0.13 -0.12 0.11 -0.048 0.038 -0.0061 0.055 0.054 -0.14 -0.27 0.92 0.11 -0.39 0.26 0.038 -0.024 -0.063 -0.026 -0.19

Imputation

1.00

- 0.75

- 0.50

- 0.25

- 0.00

- -0.25

-0.50

- -0.75

∢ זמן הריצה הוא דבר קריטי ותכונות בעלות קורלציה גבוהה תורמות מידע דומה למודל ולכן ניתן להשמיט את חלקן.

בחרנו להשמיט את: tmax, tmin, dewpoit, wetbulb, heat

Values of key_sum 5000 4000 key sum values 3000 2000 1000 4000 6000 8000 10000 12000 2000 dataframe indexes

Imputation

- במהלך ההשוואה בין
 ביצועי המודלים התגלתה
 תצפית חריגה בנתוני
 htesta.
- החלטנו להוריד את תצפית זו, כיוון שהיא לא משקפת את הרוב המוחלט של הנתונים ולכן תפגע בהערכה נכונה של דיוק המודלים.

Comparison Between True key_sum Values to Predicted True key_sum 700 KNN key_sum Regression Tree key sum 600 500 key_sum values 200 100 2000 6000 8000 10000 12000 test dataframe indexes MSE run time(seconds)

KNN Regressor 2045.002

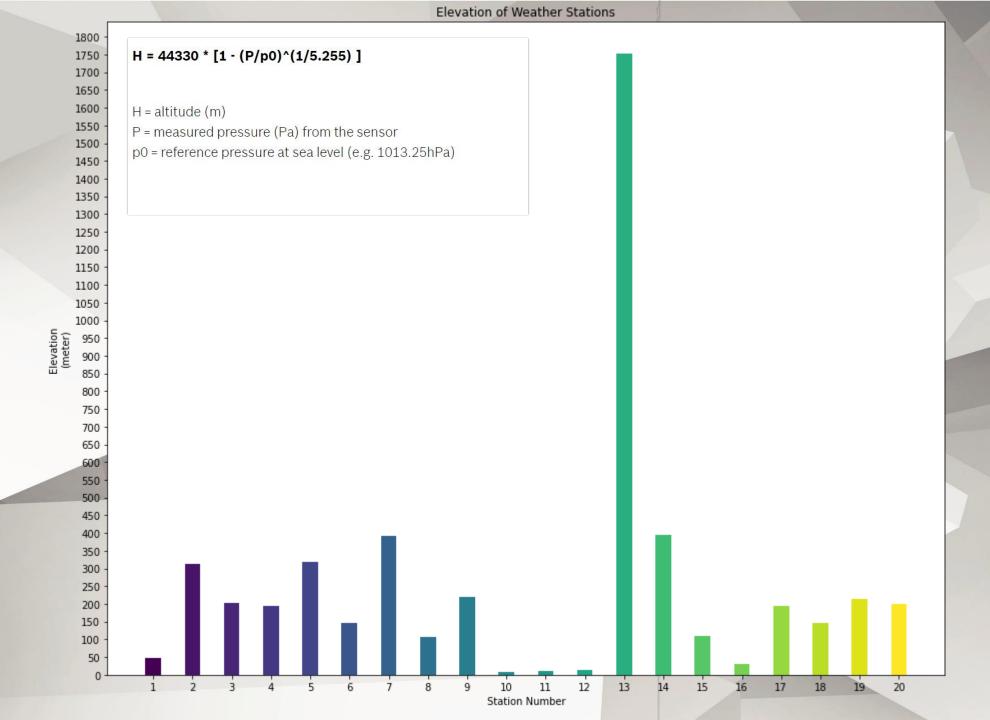
Decision Tree Regressor 1629.083

170.504590

99.390431

הערכת מודלים

- על מנת להשוות
 בין המודלים
 השתמשנו בחלק
 C בMSE C
 ריצה ובחלק
 במידת דיוק וזמן
 ריצה.
- יתן לראות את ההשוואה בין ערכי ההשוואה בין ערכי key_sum האמיתיים בסט המבחן לבין הערכים שהמודלים העריכו לגבי כל תצפית.



הערכת גובה

על מנת להעריך ♦ את הגובה של כל תחנת מזג אוויר השתמשנו במשוואה מטאורולוגית אשר משתמשת בלחץ האוויר בגובה פני הים ולחץ האוויר הנמדד בכל תחנה