FeatureAddition-MachineLearning

January 12, 2021

0.1 Select 10 random house from filtered dataset

Filtered dataset:[37, 40, 41, 42, 51, 53, 54,55, 56, 57, 58, 60, 70, 72, 99, 100, 105, 108, 115]

Based on older notebooks and on Amin's adaptation of that

[3]:
```
import pandas as pd
import numpy as np
from IPython.display import display, HTML
```

[4]:
```
df = pd.read_pickle('W12_KNMI_complete')
irr = df[['sun_irradiance']]
sun = irr.shift(24)
sun
```

[4]:
```
                     sun_irradiance
DateTime
2018-12-01 01:00:00             NaN
2018-12-01 02:00:00             NaN
2018-12-01 03:00:00             NaN
2018-12-01 04:00:00             NaN
2018-12-01 05:00:00             NaN
…                               …
2020-11-24 20:00:00             0.0
2020-11-24 21:00:00             0.0
2020-11-24 22:00:00             0.0
2020-11-24 23:00:00             0.0
2020-11-25 00:00:00             0.0

[17400 rows x 1 columns]
```

0.1.1 Features

[5]:
```
houses = [28, 37, 40, 42, 51, 56, 58, 70, 99, 100, 105, 115]
notsheet = 'consumption'
sheet = 'production'
path = '/home/16095065/notebooks/zero/Data:/testDataFrames/TEST/'
```

[11]:
```
for housenr in houses:
    housenr = f'{housenr:03}'
    df = pd.read_pickle(f'consumption_production_pickle_{housenr}')
    df = df.drop([notsheet], axis=1)

    #HOURS
    hrdf = pd.DataFrame(df.index.hour, index=df.index, columns=['hour'])
    hrdum = pd.get_dummies(hrdf['hour'], 'hour')
    df = pd.concat([df, hrdum], axis=1)


    #WEEKDAY
    wddf = pd.DataFrame(df.index.dayofweek, index=df.index, columns=['weekday'])
    wddum = pd.get_dummies(wddf['weekday'], 'weekday')
    df = pd.concat([df, wddum], axis=1)

    #Shifted days
    def shifting(sf, shift:int):
        sf['cons_T-'+str(shift)] = sf[sheet].shift(periods=shift, freq='H')
        return sf

    shiftDagen = range(24,169)
    for i in shiftDagen:
        df = shifting(df, i)

    #Mean values
    temp_df = df.filter(items=[sheet]).resample('7D').mean()
    day_temp_df = df.filter(items=[sheet]).resample('1D').mean()

        #week
    temp_df = temp_df.shift(1)
    temp_df = temp_df.resample('1H').ffill()

        #day
    day_temp_df = day_temp_df.shift(1)
    day_temp_df = day_temp_df.resample('1H').ffill()

        #columns added
    df['day_mean'] = day_temp_df#.mean(axis=1, skipna=True)
    df['week_mean'] = temp_df#.mean(axis=1, skipna=True)
    df['irradiance'] = sun

    df = df.fillna(0)
    display(HTML(f'<h3> {housenr} </h3>'))

    cols = df.columns.tolist()
    cols = cols[1:] + cols[:1]
```

```
    df = df[cols]

    display(df.head(3))
    df.to_pickle(f"{path}MachineLearning_{sheet}_{housenr}_2")
```

```
<IPython.core.display.HTML object>
```

	hour_0	hour_1	hour_2	hour_3	hour_4	hour_5	hour_6	\
2018-12-31 23:00:00	0	0	0	0	0	0	0	
2019-01-01 00:00:00	1	0	0	0	0	0	0	
2019-01-01 01:00:00	0	1	0	0	0	0	0	

	hour_7	hour_8	hour_9	…	cons_T-163	cons_T-164	\
2018-12-31 23:00:00	0	0	0	…	0.0	0.0	
2019-01-01 00:00:00	0	0	0	…	0.0	0.0	
2019-01-01 01:00:00	0	0	0	…	0.0	0.0	

	cons_T-165	cons_T-166	cons_T-167	cons_T-168	day_mean	\
2018-12-31 23:00:00	0.0	0.0	0.0	0.0	0.0	
2019-01-01 00:00:00	0.0	0.0	0.0	0.0	0.0	
2019-01-01 01:00:00	0.0	0.0	0.0	0.0	0.0	

	week_mean	irradiance	production
2018-12-31 23:00:00	0.0	0.0	0.0
2019-01-01 00:00:00	0.0	0.0	0.0
2019-01-01 01:00:00	0.0	0.0	0.0

```
[3 rows x 180 columns]
```

```
<IPython.core.display.HTML object>
```

	hour_0	hour_1	hour_2	hour_3	hour_4	hour_5	hour_6	\
2018-12-31 23:00:00	0	0	0	0	0	0	0	
2019-01-01 00:00:00	1	0	0	0	0	0	0	
2019-01-01 01:00:00	0	1	0	0	0	0	0	

	hour_7	hour_8	hour_9	…	cons_T-163	cons_T-164	\
2018-12-31 23:00:00	0	0	0	…	0.0	0.0	
2019-01-01 00:00:00	0	0	0	…	0.0	0.0	
2019-01-01 01:00:00	0	0	0	…	0.0	0.0	

	cons_T-165	cons_T-166	cons_T-167	cons_T-168	day_mean	\
2018-12-31 23:00:00	0.0	0.0	0.0	0.0	0.0	
2019-01-01 00:00:00	0.0	0.0	0.0	0.0	0.0	
2019-01-01 01:00:00	0.0	0.0	0.0	0.0	0.0	

	week_mean	irradiance	production
2018-12-31 23:00:00	0.0	0.0	0.0
2019-01-01 00:00:00	0.0	0.0	0.0

```
2019-01-01 01:00:00        0.0         0.0         0.0

[3 rows x 180 columns]

<IPython.core.display.HTML object>

                     hour_0  hour_1  hour_2  hour_3  hour_4  hour_5  hour_6  \
2018-12-31 23:00:00       0       0       0       0       0       0       0
2019-01-01 00:00:00       1       0       0       0       0       0       0
2019-01-01 01:00:00       0       1       0       0       0       0       0

                     hour_7  hour_8  hour_9  …  cons_T-163  cons_T-164  \
2018-12-31 23:00:00       0       0       0  …         0.0         0.0
2019-01-01 00:00:00       0       0       0  …         0.0         0.0
2019-01-01 01:00:00       0       0       0  …         0.0         0.0

                     cons_T-165  cons_T-166  cons_T-167  cons_T-168  day_mean  \
2018-12-31 23:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 00:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 01:00:00         0.0         0.0         0.0         0.0       0.0

                     week_mean  irradiance  production
2018-12-31 23:00:00        0.0         0.0         0.0
2019-01-01 00:00:00        0.0         0.0         0.0
2019-01-01 01:00:00        0.0         0.0         0.0

[3 rows x 180 columns]

<IPython.core.display.HTML object>

                     hour_0  hour_1  hour_2  hour_3  hour_4  hour_5  hour_6  \
2018-12-31 23:00:00       0       0       0       0       0       0       0
2019-01-01 00:00:00       1       0       0       0       0       0       0
2019-01-01 01:00:00       0       1       0       0       0       0       0

                     hour_7  hour_8  hour_9  …  cons_T-163  cons_T-164  \
2018-12-31 23:00:00       0       0       0  …         0.0         0.0
2019-01-01 00:00:00       0       0       0  …         0.0         0.0
2019-01-01 01:00:00       0       0       0  …         0.0         0.0

                     cons_T-165  cons_T-166  cons_T-167  cons_T-168  day_mean  \
2018-12-31 23:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 00:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 01:00:00         0.0         0.0         0.0         0.0       0.0

                     week_mean  irradiance  production
2018-12-31 23:00:00        0.0         0.0         0.0
2019-01-01 00:00:00        0.0         0.0         0.0
2019-01-01 01:00:00        0.0         0.0         0.0
```

```
[3 rows x 180 columns]
```

```
<IPython.core.display.HTML object>
```

```
                     hour_0  hour_1  hour_2  hour_3  hour_4  hour_5  hour_6  \
2018-12-31 23:00:00       0       0       0       0       0       0       0
2019-01-01 00:00:00       1       0       0       0       0       0       0
2019-01-01 01:00:00       0       1       0       0       0       0       0

                     hour_7  hour_8  hour_9  …  cons_T-163  cons_T-164  \
2018-12-31 23:00:00       0       0       0  …         0.0         0.0
2019-01-01 00:00:00       0       0       0  …         0.0         0.0
2019-01-01 01:00:00       0       0       0  …         0.0         0.0

                     cons_T-165  cons_T-166  cons_T-167  cons_T-168  day_mean  \
2018-12-31 23:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 00:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 01:00:00         0.0         0.0         0.0         0.0       0.0

                     week_mean  irradiance  production
2018-12-31 23:00:00        0.0         0.0         0.0
2019-01-01 00:00:00        0.0         0.0         0.0
2019-01-01 01:00:00        0.0         0.0         0.0
```

```
[3 rows x 180 columns]
```

```
<IPython.core.display.HTML object>
```

```
                     hour_0  hour_1  hour_2  hour_3  hour_4  hour_5  hour_6  \
2018-12-31 23:00:00       0       0       0       0       0       0       0
2019-01-01 00:00:00       1       0       0       0       0       0       0
2019-01-01 01:00:00       0       1       0       0       0       0       0

                     hour_7  hour_8  hour_9  …  cons_T-163  cons_T-164  \
2018-12-31 23:00:00       0       0       0  …         0.0         0.0
2019-01-01 00:00:00       0       0       0  …         0.0         0.0
2019-01-01 01:00:00       0       0       0  …         0.0         0.0

                     cons_T-165  cons_T-166  cons_T-167  cons_T-168  day_mean  \
2018-12-31 23:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 00:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 01:00:00         0.0         0.0         0.0         0.0       0.0

                     week_mean  irradiance  production
2018-12-31 23:00:00        0.0         0.0         0.0
2019-01-01 00:00:00        0.0         0.0         0.0
2019-01-01 01:00:00        0.0         0.0         0.0
```

```
[3 rows x 180 columns]
```

<IPython.core.display.HTML object>

	hour_0	hour_1	hour_2	hour_3	hour_4	hour_5	hour_6	\
2018-12-31 23:00:00	0	0	0	0	0	0	0	
2019-01-01 00:00:00	1	0	0	0	0	0	0	
2019-01-01 01:00:00	0	1	0	0	0	0	0	

	hour_7	hour_8	hour_9	…	cons_T-163	cons_T-164	\
2018-12-31 23:00:00	0	0	0	…	0.0	0.0	
2019-01-01 00:00:00	0	0	0	…	0.0	0.0	
2019-01-01 01:00:00	0	0	0	…	0.0	0.0	

	cons_T-165	cons_T-166	cons_T-167	cons_T-168	day_mean	\
2018-12-31 23:00:00	0.0	0.0	0.0	0.0	0.0	
2019-01-01 00:00:00	0.0	0.0	0.0	0.0	0.0	
2019-01-01 01:00:00	0.0	0.0	0.0	0.0	0.0	

	week_mean	irradiance	production
2018-12-31 23:00:00	0.0	0.0	0.0
2019-01-01 00:00:00	0.0	0.0	0.0
2019-01-01 01:00:00	0.0	0.0	0.0

[3 rows x 180 columns]

<IPython.core.display.HTML object>

	hour_0	hour_1	hour_2	hour_3	hour_4	hour_5	hour_6	\
2018-12-31 23:00:00	0	0	0	0	0	0	0	
2019-01-01 00:00:00	1	0	0	0	0	0	0	
2019-01-01 01:00:00	0	1	0	0	0	0	0	

	hour_7	hour_8	hour_9	…	cons_T-163	cons_T-164	\
2018-12-31 23:00:00	0	0	0	…	0.0	0.0	
2019-01-01 00:00:00	0	0	0	…	0.0	0.0	
2019-01-01 01:00:00	0	0	0	…	0.0	0.0	

	cons_T-165	cons_T-166	cons_T-167	cons_T-168	day_mean	\
2018-12-31 23:00:00	0.0	0.0	0.0	0.0	0.0	
2019-01-01 00:00:00	0.0	0.0	0.0	0.0	0.0	
2019-01-01 01:00:00	0.0	0.0	0.0	0.0	0.0	

	week_mean	irradiance	production
2018-12-31 23:00:00	0.0	0.0	0.0
2019-01-01 00:00:00	0.0	0.0	0.0
2019-01-01 01:00:00	0.0	0.0	0.0

[3 rows x 180 columns]

<IPython.core.display.HTML object>

```
                     hour_0  hour_1  hour_2  hour_3  hour_4  hour_5  hour_6  \
2018-12-31 23:00:00       0       0       0       0       0       0       0
2019-01-01 00:00:00       1       0       0       0       0       0       0
2019-01-01 01:00:00       0       1       0       0       0       0       0

                     hour_7  hour_8  hour_9  …  cons_T-163  cons_T-164  \
2018-12-31 23:00:00       0       0       0  …         0.0         0.0
2019-01-01 00:00:00       0       0       0  …         0.0         0.0
2019-01-01 01:00:00       0       0       0  …         0.0         0.0

                     cons_T-165  cons_T-166  cons_T-167  cons_T-168  day_mean  \
2018-12-31 23:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 00:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 01:00:00         0.0         0.0         0.0         0.0       0.0

                     week_mean  irradiance  production
2018-12-31 23:00:00        0.0         0.0         0.0
2019-01-01 00:00:00        0.0         0.0         0.0
2019-01-01 01:00:00        0.0         0.0         0.0

[3 rows x 180 columns]
```

```
<IPython.core.display.HTML object>
```

```
                     hour_0  hour_1  hour_2  hour_3  hour_4  hour_5  hour_6  \
2018-12-31 23:00:00       0       0       0       0       0       0       0
2019-01-01 00:00:00       1       0       0       0       0       0       0
2019-01-01 01:00:00       0       1       0       0       0       0       0

                     hour_7  hour_8  hour_9  …  cons_T-163  cons_T-164  \
2018-12-31 23:00:00       0       0       0  …         0.0         0.0
2019-01-01 00:00:00       0       0       0  …         0.0         0.0
2019-01-01 01:00:00       0       0       0  …         0.0         0.0

                     cons_T-165  cons_T-166  cons_T-167  cons_T-168  day_mean  \
2018-12-31 23:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 00:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 01:00:00         0.0         0.0         0.0         0.0       0.0

                     week_mean  irradiance  production
2018-12-31 23:00:00        0.0         0.0         0.0
2019-01-01 00:00:00        0.0         0.0         0.0
2019-01-01 01:00:00        0.0         0.0         0.0

[3 rows x 180 columns]
```

```
<IPython.core.display.HTML object>
```

```
                     hour_0  hour_1  hour_2  hour_3  hour_4  hour_5  hour_6  \
2018-12-31 23:00:00       0       0       0       0       0       0       0
```

```
2019-01-01 00:00:00       1       0       0       0       0       0       0
2019-01-01 01:00:00       0       1       0       0       0       0       0

                     hour_7  hour_8  hour_9  …  cons_T-163  cons_T-164  \
2018-12-31 23:00:00       0       0       0  …         0.0         0.0
2019-01-01 00:00:00       0       0       0  …         0.0         0.0
2019-01-01 01:00:00       0       0       0  …         0.0         0.0

                     cons_T-165  cons_T-166  cons_T-167  cons_T-168  day_mean  \
2018-12-31 23:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 00:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 01:00:00         0.0         0.0         0.0         0.0       0.0

                     week_mean  irradiance  production
2018-12-31 23:00:00        0.0         0.0         0.0
2019-01-01 00:00:00        0.0         0.0         0.0
2019-01-01 01:00:00        0.0         0.0         0.0

[3 rows x 180 columns]
```

```
<IPython.core.display.HTML object>
```

```
                     hour_0  hour_1  hour_2  hour_3  hour_4  hour_5  hour_6  \
2018-12-31 23:00:00       0       0       0       0       0       0       0
2019-01-01 00:00:00       1       0       0       0       0       0       0
2019-01-01 01:00:00       0       1       0       0       0       0       0

                     hour_7  hour_8  hour_9  …  cons_T-163  cons_T-164  \
2018-12-31 23:00:00       0       0       0  …         0.0         0.0
2019-01-01 00:00:00       0       0       0  …         0.0         0.0
2019-01-01 01:00:00       0       0       0  …         0.0         0.0

                     cons_T-165  cons_T-166  cons_T-167  cons_T-168  day_mean  \
2018-12-31 23:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 00:00:00         0.0         0.0         0.0         0.0       0.0
2019-01-01 01:00:00         0.0         0.0         0.0         0.0       0.0

                     week_mean  irradiance  production
2018-12-31 23:00:00        0.0         0.0         0.0
2019-01-01 00:00:00        0.0         0.0         0.0
2019-01-01 01:00:00        0.0         0.0         0.0

[3 rows x 180 columns]
```

[ ]: