Windfinder

A basic machine learning approach on predicting the pageviews for the windfinder website when provided with weatherdata.

Participants:

Kilian Köln

Course and Semester

Deep Learning from Scratch, SoSe2021

License

Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Introduction

The website https://www.windfinder.com provides weather data which can be used for different hobbies like kiting or surfing to check if the conditions are right. The page views of the website is fluctuating. It seems that, when the conditions are good for surfing or kiting, the page views tend to be higher. Thus, the page views could be interpreted as a proxy for how good a certain weather condition is for those activities.

With this work, an attempt is made to predict the page views from the weather data. Past page views and weather data is used to train a model, which then could be used to predict new page views from given weather data.

Data and Methods

Hourly weatherdata from windfinder from 2014 to 2019 is used and group to daily averages. This data is then matched with daily pageviews which for the windfinder website. From the weatherdata temperature, page views, air temperature and location are feed into the model to predict the pagviews.

Results

The mode predicts the page views with an accuracy of around 80 %.

Baseline

A basic linear regression achieves an accuracy of 68 %, therefore the neural network model is clearly better.

```
In [6]:
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
         import glob
         import tensorflow as tf
         from sklearn.model_selection import train_test_split
         from sklearn.preprocessing import StandardScaler
         from sklearn import linear_model
         from sklearn.metrics import mean absolute error
         import keras
         from keras.models import Sequential
         from keras.layers import Dense
         %matplotlib inline
         print(tf.version)
```

<module 'tensorflow._api.v2.version' from '/home/kilian/.local/lib/python3.
8/site-packages/tensorflow/_api/v2/version/__init__.py'>

Loading the data

To begin with, the data is loaded from the files and put into pandas data frames. In this case only the data for Kiel, St. Peter-Ording and Warnemünde is considered. The code below will only work if the data itself is provided in the folders, which it is not for copyright considerations of the data from wind finder.

The data itself is cannot be provided alongside with this notebook, because of copyright considerations of wind finder. For this notebook to work, the corresponding data needs to be put in the folder structure.

```
In [5]:
         pageviews_kiel = pd.concat([pd.read_csv(f, sep = ';',names = ["date", "page")
                                     for f in glob.glob('data windfinder/SelectedLog
         pageviews sktPeterOrding = pd.concat([pd.read csv(f, sep = ';',names = ["data"
                                     for f in glob.glob('data_windfinder/SelectedLog
         pageviews_warnemuende = pd.concat([pd.read_csv(f, sep = ';',names = ["date")]
                                     for f in glob.glob('data_windfinder/SelectedLog
         weatherdata kiel = pd.concat([pd.read csv(f, sep = ';')
                                      for f in glob.glob('data windfinder/SelectedLog
         weatherdata_kiel.columns=["date", "wind", "wind direction", "wind gust", "
         weatherdata kiel = weatherdata kiel.assign(location = 1)
         weatherdata_sktPeterOrding = pd.concat([pd.read_csv(f, sep = ';')
                                     for f in glob.glob('data_windfinder/SelectedLog
         weatherdata_sktPeterOrding.columns=["date", "wind", "wind direction", "wind"]
         weatherdata sktPeterOrding = weatherdata sktPeterOrding.assign(location =
         weatherdata warnemuende = pd.concat([pd.read csv(f, sep = ';')
                                     for f in glob.glob('data_windfinder/SelectedLog
         weatherdata_warnemuende.columns=["date", "wind", "wind direction", "wind g
         weatherdata warnemuende = weatherdata warnemuende.assign(location = 3)
```

In [220...

```
print(pageviews_kiel.head(5))
print(weatherdata_kiel.head(5))
```

```
date pageviews
 20190101
                 2954
  20190102
                 2304
2
  20190103
                 1225
3 20190104
                 1288
  20190105
                  840
                 date wind wind direction wind gust air temp location
0 2016-01-01 00:08:00 16.0
                                        208
                                                  17.0
                                                             6.0
                                                                         1
  2016-01-01 01:36:00 12.0
                                        194
                                                  14.0
                                                             6.0
                                                                         1
  2016-01-01 01:52:00 14.0
                                        193
                                                  14.0
                                                             6.0
                                                                         1
                                                  14.0
  2016-01-01 02:08:00 12.0
                                        192
                                                             6.0
                                                                         1
  2016-01-01 02:16:00 12.0
                                        185
                                                  16.0
                                                                         1
                                                             6.0
```

Helper functions for data preparation

group_hourly_weatherdata_to_daily() is used to create daily weather data from the provided hourly weather data. The reasons for this, is that the pageviews are also provided daily and they need to be matched.

prepare_pageviews() is used to convert the integer number which is used as a date in the default date format as used in pandas.

```
In [24]: def group_hourly_weatherdata_to_daily(weatherdata):
    weatherdata["date"] = pd.to_datetime(weatherdata["date"])
    weatherdata = weatherdata.resample('D', on="date").mean()
    weatherdata = weatherdata.reset_index()
    return weatherdata
```

Preparing the data

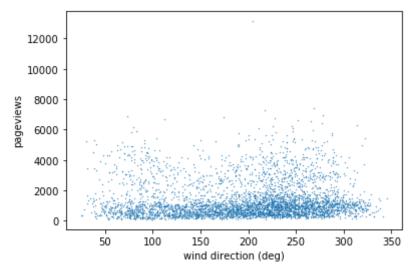
The data is combined and rows with missing entries are deleted.

```
In [27]:
          #preparing data
          pageviews_kiel = prepare_pageviews(pageviews_kiel)
          pageviews_sktPeterOrding = prepare_pageviews(pageviews_sktPeterOrding)
          pageviews_warnemuende = prepare_pageviews(pageviews_warnemuende)
          #group hourly data to daily data
          daily_weather_kiel = group_hourly_weatherdata_to_daily(weatherdata_kiel)
          daily_weather_sktPeterOrding = group_hourly_weatherdata_to_daily(weatherdat
          daily weather warnemuende = group hourly weatherdata to daily(weatherdata \
          #combine pageviews and weatherdata into one data frame
          kiel_combined = pd.merge(daily_weather_kiel,pageviews_kiel, on = 'date')
          sktPeterOrding_combined = pd.merge(daily_weather_sktPeterOrding,pageviews_
          warnemuende combined = pd merge(daily weather warnemuende,pageviews warnem
          combined data = pd.concat([kiel combined, sktPeterOrding combined,warnemuel
          combined data = combined data.reset index(drop="True")
          print(combined_data)
In [30]:
          # check for missing values
          print("missing values in data: ", combined_data.isnull().sum())
                                                      0
         missing values in data: date
         wind
                           52
                           52
         wind direction
                           83
         wind gust
         air temp
                           51
         location
                           51
         pageviews
         dtype: int64
In [212...
          #seperate rows with missing data from dataset
          incomplete rows = combined data[combined data.isna().any(axis=1)]
          combined data = combined data.dropna()
          print("missing values in data: ", combined_data.isnull().sum())
         missing values in data: date
                                                     0
         wind
         wind direction
                           0
                           0
         wind gust
         air temp
                           0
                           0
         location
                           0
         pageviews
```

dtvne: int64

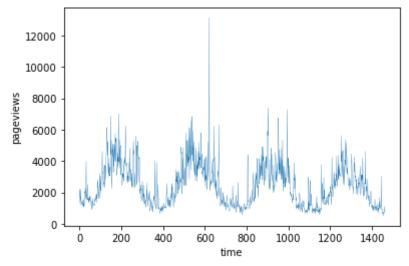
Data exploration

```
In [51]:
           wind = combined_data.loc[:,'wind']
           views = combined_data.loc[:,'pageviews']
           plt.scatter(wind, views, s=0.1)
           plt.xlabel("wind(kts)")
           plt.ylabel("pageviews")
           plt.show()
            12000
            10000
             8000
          pageviews
             6000
             4000
             2000
                0
                           10
                                 15
                                            25
                                                 30
                                      20
                                                       35
                                                            40
                                      wind(kts)
In [56]:
           temp = combined_data.loc[:,'air temp']
           plt.scatter(temp, views, s=0.1)
           plt.xlabel("temperature (C°)")
           plt.ylabel("pageviews")
           plt.show()
            12000
            10000
             8000
             6000
             4000
             2000
                                          10
                                               15
                                                     20
                                                           25
                  -10
                                   temperature (C°)
In [58]:
           wind_direction = combined_data.loc[:,'wind direction']
           plt.scatter(wind_direction, views, s=0.1)
           plt.xlabel("wind direction (deg)")
           plt.ylabel("pageviews")
           plt.show()
```



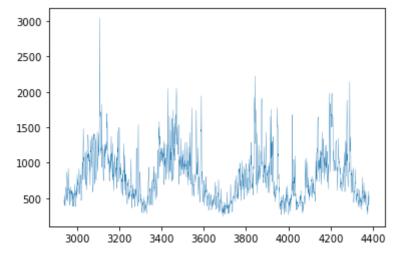
```
In [216... data_kiel = combined_data[combined_data['location'] == 1]
    data_sktPeterOrding = combined_data[combined_data['location'] == 2]
    data_warnemuende = combined_data[combined_data['location'] == 3]

data_kiel['pageviews'].plot(linewidth=0.3, xlabel = "time", ylabel = "page")
```

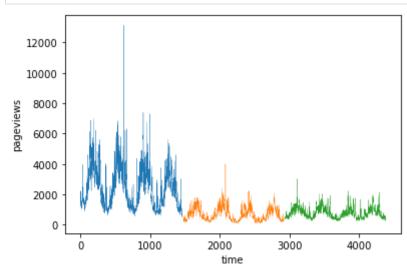


```
In [67]:
            data_sktPeterOrding['pageviews'].plot(linewidth=0.3, xlabel = "time", ylabe
           4000
           3500
           3000
           2500
           2000
           1500
           1000
            500
              0
                     1600
                           1800
                                  2000
                                        2200
                                              2400
                                                     2600
                                                           2800
```

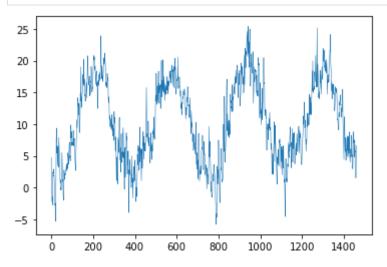
In [68]: data_warnemuende['pageviews'].plot(linewidth=0.3, xlabel = "time", ylabel



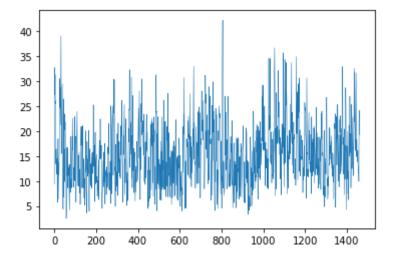
In [218... data_kiel['pageviews'].plot(linewidth=0.3, xlabel = "time", ylabel = "pageviews'].plot(linewidth=0.3, xlabel = "time", ylabel data_warnemuende['pageviews'].plot(linewidth=0.3, xlabel = "time", ylabel = "time", y



In [71]: data_kiel['air temp'].plot(linewidth=0.5);



```
In [72]: data_kiel['wind'].plot(linewidth=0.5);
```



- contradicting my expectations, the popularity of the wind finder website seems to be stable over the 4 years for which I have data.
- as expected, both the wind speed and the temperature are positively correlated with the number of page views
- the repeating sin-like patter of the page views seems to match the repeating pattern of the air temperature. I would assume that the most important factor for the page views is the air temperature.
- the wind direction seems not to be very important for the prediction of page views

Splitting the Data and building the model

```
X = combined data.drop(columns=['pageviews','date', 'wind direction','wind
In [192...
          y = combined data.drop(columns=['date', 'wind', 'wind direction', 'wind gu
          X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33,
In [193...
          print(X_train.describe())
                        wind
                                  air temp
                                                location
                 2881.000000
                               2881.000000
                                             2881.000000
          count
                   12.240006
                                 10.527809
          mean
                                                2.012149
                                                0.814916
          std
                    5.846952
                                  6.456361
          min
                    2.956522
                                 -9.000000
                                                1.000000
          25%
                    7.583333
                                  5.384615
                                                1.000000
          50%
                   11.041667
                                 10.208333
                                                2.000000
          75%
                   15.833333
                                 16.152778
                                                3.000000
                   41.671329
                                 27.708333
                                                3.000000
          max
In [194...
          print(X_test.describe())
                        wind
                                  air temp
                                                location
                 1419.000000
                               1419.000000
                                             1419.000000
          count
                   12.665607
                                 10.104142
                                                1.970402
          mean
          std
                    5.996547
                                  6.329566
                                                0.819696
                                 -8.291667
                                                1.000000
          min
                    2.513274
          25%
                    8.031944
                                  5.272727
                                                1.000000
          50%
                   11.416667
                                  9.541667
                                                2.000000
          75%
                   16.104286
                                 15.750000
                                                3.000000
```

8 of 22 7/31/21, 02:19

3.000000

26.333333

42.244755

max

```
In [195... # scaling the data
       sc = StandardScaler()
       X_train = sc.fit_transform(X_train)
       X test = sc.transform(X test)
      # building the model
In [209...
       model = Sequential()
       d rate = 0.8
       reg = 0
       model.add(Dense(256, activation='relu', kernel_regularizer=regularizers.l2(
       Dropout(d rate)
       model.add(Dense(128,activation='relu', kernel regularizer=regularizers.l2(re
       Dropout(d rate)
       model.add(Dense(128,activation='relu', kernel regularizer=regularizers.l2(re
       Dropout(d rate)
       model.add(Dense(16,activation='relu',kernel_regularizer=regularizers.l2(regularizers.l2)
       model.add(Dense(1, activation='relu'))
       # Compile model
       model.compile(loss='mae', optimizer='adam', metrics=['mae'])
       # Fit the model
       history = model.fit(X_train, y_train, epochs=200, batch_size=100,
                     validation data=(X test, y test),
                    verbose=1)
       # evaluate the model
       scores = model.evaluate(X_test, y_test, verbose=0)
       print("%s:" % (model.metrics names[1]))
      Epoch 1/200
      e: 1321.8330 - val loss: 1312.9773 - val mae: 1312.9774
      Epoch 2/200
      e: 1241.4580 - val loss: 1044.0126 - val mae: 1044.0126
      Epoch 3/200
      e: 898.9895 - val loss: 570.6832 - val mae: 570.6832
      Epoch 4/200
      e: 541.5541 - val_loss: 457.1435 - val_mae: 457.1435
      Epoch 5/200
      e: 445.8793 - val loss: 414.3849 - val mae: 414.3849
      e: 423.9984 - val loss: 394.9117 - val mae: 394.9117
      Epoch 7/200
      e: 395.3664 - val_loss: 385.5625 - val_mae: 385.5625
      e: 393.5070 - val loss: 375.4124 - val mae: 375.4124
      Epoch 9/200
      e: 372.5546 - val loss: 369.1357 - val mae: 369.1357
      Epoch 10/200
      e: 373.4144 - val loss: 366.2620 - val mae: 366.2620
```

```
Epoch 11/200
e: 361.5205 - val_loss: 360.6183 - val_mae: 360.6183
Epoch 12/200
e: 375.1317 - val_loss: 357.8358 - val_mae: 357.8358
Epoch 13/200
e: 366.6812 - val loss: 353.6728 - val mae: 353.6728
Epoch 14/200
e: 362.7484 - val_loss: 350.1714 - val_mae: 350.1714
Epoch 15/200
e: 359.5532 - val_loss: 349.8632 - val_mae: 349.8632
Epoch 16/200
e: 365.1108 - val loss: 347.1161 - val mae: 347.1161
Epoch 17/200
e: 351.9534 - val loss: 342.2820 - val mae: 342.2820
Epoch 18/200
e: 360.0468 - val loss: 339.4390 - val mae: 339.4390
Epoch 19/200
e: 349.7781 - val loss: 337.0977 - val mae: 337.0977
Epoch 20/200
e: 351.5255 - val loss: 344.0683 - val mae: 344.0683
Epoch 21/200
e: 347.0518 - val loss: 336.2761 - val mae: 336.2761
Epoch 22/200
e: 341.8795 - val loss: 332.3392 - val mae: 332.3392
Epoch 23/200
29/29 [============== ] - Os 2ms/step - loss: 346.9397 - ma
e: 346.9396 - val loss: 331.4354 - val mae: 331.4354
Epoch 24/200
e: 333.6759 - val loss: 329.6468 - val mae: 329.6468
Epoch 25/200
e: 346.6793 - val_loss: 327.4880 - val_mae: 327.4880
Epoch 26/200
e: 344.4411 - val loss: 329.1262 - val mae: 329.1262
Epoch 27/200
e: 326.9577 - val loss: 325.8958 - val mae: 325.8958
Epoch 28/200
e: 341.0396 - val loss: 326.6912 - val mae: 326.6912
e: 325.8117 - val loss: 326.3834 - val mae: 326.3834
Epoch 30/200
e: 332.7764 - val loss: 322.9972 - val mae: 322.9972
Epoch 31/200
e: 341.7166 - val_loss: 322.2048 - val_mae: 322.2048
Epoch 32/200
```

```
e: 337.9061 - val_loss: 322.1075 - val_mae: 322.1075
Epoch 33/200
29/29 [=============== ] - 0s 2ms/step - loss: 331.3986 - ma
e: 331.3986 - val_loss: 323.2071 - val_mae: 323.2071
Epoch 34/200
29/29 [============== ] - Os 2ms/step - loss: 333.0525 - ma
e: 333.0525 - val_loss: 324.1192 - val_mae: 324.1192
Epoch 35/200
e: 333.5655 - val_loss: 321.2472 - val_mae: 321.2472
Epoch 36/200
e: 326.4668 - val_loss: 320.6242 - val_mae: 320.6242
Epoch 37/200
e: 318.0150 - val_loss: 319.8421 - val_mae: 319.8421
Epoch 38/200
29/29 [============== ] - Os 2ms/step - loss: 335.4658 - ma
e: 335.4658 - val loss: 321.2374 - val mae: 321.2374
Epoch 39/200
e: 325.7424 - val loss: 330.6827 - val mae: 330.6826
Epoch 40/200
29/29 [============== ] - Os 2ms/step - loss: 326.7814 - ma
e: 326.7814 - val_loss: 320.3814 - val_mae: 320.3814
Epoch 41/200
e: 323.5118 - val loss: 319.7118 - val mae: 319.7118
Epoch 42/200
29/29 [============= ] - Os 2ms/step - loss: 324.9710 - ma
e: 324.9710 - val loss: 318.6366 - val mae: 318.6366
Epoch 43/200
e: 325.1526 - val loss: 318.9670 - val mae: 318.9670
Epoch 44/200
e: 328.8987 - val loss: 318.0964 - val mae: 318.0964
Epoch 45/200
e: 346.5080 - val loss: 318.0948 - val mae: 318.0948
Epoch 46/200
29/29 [============= ] - Os 2ms/step - loss: 330.1352 - ma
e: 330.1352 - val loss: 318.5973 - val mae: 318.5973
Epoch 47/200
e: 317.2659 - val loss: 317.0535 - val mae: 317.0535
Epoch 48/200
e: 323.7609 - val loss: 318.7151 - val mae: 318.7152
e: 333.8454 - val loss: 316.1389 - val mae: 316.1389
Epoch 50/200
e: 322.0183 - val_loss: 317.6796 - val_mae: 317.6796
Epoch 51/200
e: 329.6547 - val loss: 317.9353 - val mae: 317.9353
Epoch 52/200
e: 323.9151 - val loss: 316.4948 - val mae: 316.4948
Epoch 53/200
```

```
e: 324.5670 - val loss: 318.5711 - val mae: 318.5711
Epoch 54/200
29/29 [=============== ] - 0s 2ms/step - loss: 330.4295 - ma
e: 330.4295 - val_loss: 316.2730 - val_mae: 316.2730
Epoch 55/200
29/29 [============== ] - Os 2ms/step - loss: 320.3203 - ma
e: 320.3203 - val_loss: 315.1126 - val_mae: 315.1127
29/29 [=============== ] - 0s 2ms/step - loss: 317.4527 - ma
e: 317.4527 - val loss: 326.5543 - val mae: 326.5543
Epoch 57/200
e: 323.0927 - val_loss: 315.4386 - val_mae: 315.4386
Epoch 58/200
e: 328.1424 - val loss: 317.2000 - val mae: 317.2000
Epoch 59/200
e: 321.3136 - val loss: 315.0144 - val mae: 315.0144
Epoch 60/200
e: 312.6042 - val loss: 315.0782 - val mae: 315.0783
Epoch 61/200
e: 326.5270 - val loss: 315.1234 - val mae: 315.1234
Epoch 62/200
e: 315.2720 - val loss: 315.2585 - val mae: 315.2585
Epoch 63/200
e: 325.5984 - val loss: 315.8672 - val mae: 315.8672
Epoch 64/200
29/29 [============== ] - Os 2ms/step - loss: 327.6295 - ma
e: 327.6295 - val_loss: 316.8097 - val_mae: 316.8097
Epoch 65/200
29/29 [============= ] - Os 2ms/step - loss: 328.4208 - ma
e: 328.4208 - val loss: 313.4890 - val mae: 313.4890
Epoch 66/200
29/29 [============== ] - 0s 2ms/step - loss: 317.8282 - ma
e: 317.8282 - val loss: 313.6523 - val mae: 313.6523
Epoch 67/200
e: 315.6314 - val loss: 313.9176 - val mae: 313.9176
Epoch 68/200
29/29 [============= ] - 0s 2ms/step - loss: 320.2567 - ma
e: 320.2567 - val loss: 313.2576 - val mae: 313.2576
Epoch 69/200
e: 326.5765 - val loss: 314.4983 - val mae: 314.4983
Epoch 70/200
e: 324.1825 - val loss: 312.7388 - val mae: 312.7388
Epoch 71/200
e: 321.2176 - val loss: 312.7823 - val mae: 312.7823
Epoch 72/200
29/29 [============== ] - Os 2ms/step - loss: 321.1901 - ma
e: 321.1901 - val loss: 314.7500 - val mae: 314.7500
29/29 [============== ] - 0s 2ms/step - loss: 323.5779 - ma
e: 323.5779 - val loss: 312.6456 - val mae: 312.6456
Epoch 74/200
e: 304.8816 - val loss: 313.5128 - val mae: 313.5128
```

```
Epoch 75/200
29/29 [============== ] - 0s 2ms/step - loss: 317.2227 - ma
e: 317.2227 - val_loss: 312.0528 - val_mae: 312.0528
Epoch 76/200
e: 320.9475 - val_loss: 311.6820 - val_mae: 311.6820
Epoch 77/200
29/29 [=============== ] - 0s 2ms/step - loss: 316.1557 - ma
e: 316.1557 - val loss: 315.8517 - val mae: 315.8517
Epoch 78/200
e: 311.8714 - val_loss: 313.1668 - val_mae: 313.1668
Epoch 79/200
e: 313.2642 - val_loss: 313.3619 - val_mae: 313.3619
Epoch 80/200
e: 315.3700 - val loss: 314.0073 - val mae: 314.0073
Epoch 81/200
29/29 [============== ] - Os 2ms/step - loss: 297.1491 - ma
e: 297.1491 - val loss: 311.5197 - val mae: 311.5197
Epoch 82/200
e: 313.7235 - val loss: 312.5720 - val mae: 312.5720
Epoch 83/200
e: 311.0009 - val_loss: 312.5115 - val_mae: 312.5115
Epoch 84/200
e: 310.3665 - val loss: 318.6797 - val mae: 318.6797
Epoch 85/200
e: 319.4122 - val loss: 312.1553 - val mae: 312.1553
Epoch 86/200
e: 316.5613 - val loss: 312.2453 - val mae: 312.2453
Epoch 87/200
29/29 [============= ] - Os 2ms/step - loss: 314.6265 - ma
e: 314.6265 - val loss: 314.1346 - val mae: 314.1346
Epoch 88/200
e: 322.9356 - val loss: 313.9586 - val mae: 313.9586
Epoch 89/200
e: 318.4474 - val_loss: 312.9817 - val_mae: 312.9817
Epoch 90/200
e: 317.7864 - val_loss: 311.2268 - val mae: 311.2268
Epoch 91/200
29/29 [============== ] - 0s 2ms/step - loss: 319.2909 - ma
e: 319.2909 - val loss: 312.9185 - val mae: 312.9185
Epoch 92/200
e: 322.4619 - val loss: 315.1061 - val mae: 315.1061
e: 314.4851 - val loss: 314.6669 - val mae: 314.6669
Epoch 94/200
e: 311.6825 - val loss: 315.5038 - val mae: 315.5038
Epoch 95/200
e: 315.0759 - val_loss: 312.2886 - val_mae: 312.2886
Epoch 96/200
```

```
e: 305.4058 - val_loss: 311.6821 - val_mae: 311.6821
Epoch 97/200
e: 323.0263 - val_loss: 311.6418 - val_mae: 311.6418
Epoch 98/200
29/29 [============== ] - Os 2ms/step - loss: 298.1578 - ma
e: 298.1578 - val_loss: 311.3998 - val_mae: 311.3998
Epoch 99/200
29/29 [============== ] - 0s 2ms/step - loss: 332.4002 - ma
e: 332.4002 - val_loss: 312.9706 - val_mae: 312.9706
Epoch 100/200
e: 323.1705 - val_loss: 311.5500 - val_mae: 311.5500
Epoch 101/200
e: 312.8805 - val_loss: 311.8378 - val_mae: 311.8378
Epoch 102/200
e: 307.8414 - val loss: 312.1571 - val mae: 312.1571
Epoch 103/200
e: 311.3626 - val loss: 314.2042 - val mae: 314.2041
Epoch 104/200
e: 316.2706 - val_loss: 312.1050 - val_mae: 312.1050
Epoch 105/200
29/29 [============= ] - Os 2ms/step - loss: 317.0660 - ma
e: 317.0659 - val loss: 311.7670 - val mae: 311.7670
Epoch 106/200
e: 306.2700 - val loss: 311.8951 - val mae: 311.8951
Epoch 107/200
e: 292.9186 - val loss: 311.8120 - val mae: 311.8120
Epoch 108/200
e: 311.8847 - val_loss: 312.2263 - val_mae: 312.2263
Epoch 109/200
e: 308.3614 - val loss: 312.3921 - val mae: 312.3921
Epoch 110/200
29/29 [============== ] - Os 2ms/step - loss: 312.8383 - ma
e: 312.8383 - val loss: 311.7451 - val mae: 311.7451
Epoch 111/200
e: 313.0771 - val loss: 318.8123 - val mae: 318.8123
Epoch 112/200
e: 316.4959 - val loss: 318.6669 - val mae: 318.6669
e: 321.2065 - val loss: 315.7310 - val mae: 315.7310
Epoch 114/200
e: 318.9031 - val_loss: 312.9064 - val_mae: 312.9064
Epoch 115/200
e: 320.0401 - val loss: 312.4046 - val mae: 312.4046
Epoch 116/200
e: 321.6594 - val loss: 313.7177 - val mae: 313.7177
Epoch 117/200
```

```
e: 311.8461 - val loss: 314.6604 - val mae: 314.6604
Epoch 118/200
29/29 [=============== ] - 0s 2ms/step - loss: 306.1966 - ma
e: 306.1966 - val_loss: 311.9359 - val_mae: 311.9359
Epoch 119/200
29/29 [============== ] - Os 2ms/step - loss: 304.6532 - ma
e: 304.6532 - val_loss: 312.2290 - val_mae: 312.2290
29/29 [=============== ] - 0s 2ms/step - loss: 317.1112 - ma
e: 317.1112 - val loss: 313.0789 - val mae: 313.0789
Epoch 121/200
e: 309.8670 - val_loss: 313.1803 - val_mae: 313.1803
Epoch 122/200
e: 308.4556 - val loss: 312.9465 - val mae: 312.9465
Epoch 123/200
e: 317.3573 - val loss: 312.0298 - val mae: 312.0298
Epoch 124/200
e: 316.9791 - val loss: 312.8944 - val mae: 312.8944
Epoch 125/200
e: 312.7494 - val loss: 313.0903 - val mae: 313.0903
Epoch 126/200
e: 317.0015 - val loss: 313.2946 - val mae: 313.2946
Epoch 127/200
e: 318.4482 - val loss: 313.1601 - val mae: 313.1601
Epoch 128/200
e: 311.6035 - val_loss: 313.1756 - val_mae: 313.1756
Epoch 129/200
29/29 [============== ] - Os 2ms/step - loss: 315.1252 - ma
e: 315.1252 - val loss: 312.3178 - val mae: 312.3178
Epoch 130/200
29/29 [============== ] - 0s 2ms/step - loss: 310.2222 - ma
e: 310.2222 - val loss: 313.3598 - val mae: 313.3598
Epoch 131/200
e: 318.2928 - val loss: 313.5220 - val mae: 313.5220
Epoch 132/200
e: 308.5728 - val loss: 311.4137 - val mae: 311.4137
Epoch 133/200
e: 301.6787 - val loss: 312.4032 - val mae: 312.4032
Epoch 134/200
e: 318.1683 - val loss: 323.5783 - val mae: 323.5783
Epoch 135/200
e: 320.4320 - val loss: 312.3421 - val mae: 312.3421
Epoch 136/200
29/29 [============= ] - Os 2ms/step - loss: 312.7608 - ma
e: 312.7608 - val_loss: 310.7161 - val_mae: 310.7161
Epoch 137/200
29/29 [============== ] - 0s 2ms/step - loss: 308.1154 - ma
e: 308.1154 - val loss: 312.4596 - val mae: 312.4596
Epoch 138/200
e: 316.5467 - val loss: 312.8258 - val mae: 312.8258
```

```
Epoch 139/200
29/29 [=============== ] - 0s 2ms/step - loss: 324.6939 - ma
e: 324.6939 - val_loss: 314.3152 - val_mae: 314.3152
Epoch 140/200
e: 309.5861 - val_loss: 312.4939 - val_mae: 312.4939
Epoch 141/200
29/29 [=============== ] - 0s 2ms/step - loss: 306.9751 - ma
e: 306.9751 - val loss: 315.0424 - val mae: 315.0424
Epoch 142/200
29/29 [=============== ] - 0s 2ms/step - loss: 318.4790 - ma
e: 318.4790 - val_loss: 312.1799 - val_mae: 312.1799
Epoch 143/200
e: 317.5862 - val_loss: 313.1071 - val_mae: 313.1071
Epoch 144/200
e: 313.5496 - val loss: 312.5420 - val mae: 312.5420
Epoch 145/200
e: 327.4318 - val loss: 312.1828 - val mae: 312.1828
Epoch 146/200
e: 309.4702 - val loss: 313.9972 - val mae: 313.9972
Epoch 147/200
e: 315.0657 - val_loss: 313.5005 - val_mae: 313.5005
Epoch 148/200
e: 303.3955 - val loss: 316.8821 - val mae: 316.8821
Epoch 149/200
e: 306.5889 - val loss: 313.3159 - val mae: 313.3159
Epoch 150/200
e: 307.0544 - val loss: 313.6750 - val mae: 313.6750
Epoch 151/200
29/29 [============== ] - Os 2ms/step - loss: 306.6993 - ma
e: 306.6993 - val loss: 313.1047 - val mae: 313.1047
Epoch 152/200
e: 305.7809 - val loss: 312.7703 - val mae: 312.7703
Epoch 153/200
e: 314.9317 - val_loss: 314.0162 - val_mae: 314.0162
Epoch 154/200
e: 311.7301 - val_loss: 311.3447 - val_mae: 311.3447
Epoch 155/200
29/29 [=============== ] - 0s 2ms/step - loss: 318.1554 - ma
e: 318.1554 - val loss: 313.3901 - val mae: 313.3901
Epoch 156/200
e: 297.9330 - val loss: 313.6499 - val mae: 313.6500
Epoch 157/200
e: 316.1889 - val loss: 313.2504 - val mae: 313.2504
Epoch 158/200
e: 304.9591 - val loss: 312.6682 - val mae: 312.6682
Epoch 159/200
e: 316.3063 - val_loss: 312.0679 - val_mae: 312.0679
Epoch 160/200
```

```
e: 319.7957 - val_loss: 316.6716 - val_mae: 316.6716
Epoch 161/200
e: 307.3738 - val_loss: 312.9336 - val_mae: 312.9336
Epoch 162/200
29/29 [============== ] - Os 2ms/step - loss: 305.4614 - ma
e: 305.4614 - val_loss: 311.9902 - val_mae: 311.9902
Epoch 163/200
29/29 [=============== ] - 0s 2ms/step - loss: 317.1571 - ma
e: 317.1571 - val_loss: 315.6899 - val_mae: 315.6899
Epoch 164/200
e: 311.2756 - val_loss: 312.4877 - val_mae: 312.4877
Epoch 165/200
e: 321.4161 - val_loss: 313.0511 - val_mae: 313.0511
Epoch 166/200
e: 313.7146 - val loss: 314.4171 - val mae: 314.4171
Epoch 167/200
e: 305.9766 - val loss: 314.0941 - val mae: 314.0941
Epoch 168/200
e: 303.2224 - val_loss: 315.4606 - val_mae: 315.4606
Epoch 169/200
e: 321.1621 - val loss: 311.9155 - val mae: 311.9155
Epoch 170/200
e: 312.0436 - val loss: 312.2388 - val mae: 312.2387
Epoch 171/200
e: 298.5884 - val loss: 313.9656 - val mae: 313.9656
Epoch 172/200
e: 319.9362 - val loss: 313.7073 - val mae: 313.7073
Epoch 173/200
e: 302.8285 - val loss: 313.1014 - val mae: 313.1014
Epoch 174/200
29/29 [============= ] - Os 2ms/step - loss: 303.5065 - ma
e: 303.5065 - val loss: 315.3252 - val mae: 315.3252
Epoch 175/200
e: 314.7163 - val loss: 312.6152 - val mae: 312.6152
Epoch 176/200
e: 310.7756 - val loss: 311.2538 - val mae: 311.2538
e: 302.8976 - val loss: 313.3276 - val mae: 313.3276
Epoch 178/200
e: 322.9221 - val_loss: 313.4740 - val_mae: 313.4740
Epoch 179/200
e: 306.8902 - val loss: 314.0660 - val mae: 314.0660
Epoch 180/200
29/29 [============== ] - 0s 2ms/step - loss: 317.0709 - ma
e: 317.0709 - val loss: 314.1300 - val mae: 314.1300
Epoch 181/200
```

```
e: 301.5213 - val loss: 312.3725 - val mae: 312.3725
Epoch 182/200
29/29 [============== ] - 0s 2ms/step - loss: 305.3459 - ma
e: 305.3459 - val_loss: 313.3723 - val_mae: 313.3723
Epoch 183/200
e: 314.0921 - val_loss: 314.3200 - val_mae: 314.3200
29/29 [=============== ] - 0s 2ms/step - loss: 319.1175 - ma
e: 319.1175 - val loss: 316.3937 - val mae: 316.3937
Epoch 185/200
e: 299.1478 - val_loss: 312.4559 - val_mae: 312.4559
Epoch 186/200
e: 304.4889 - val loss: 313.5106 - val mae: 313.5106
Epoch 187/200
29/29 [============== ] - Os 2ms/step - loss: 306.5190 - ma
e: 306.5190 - val loss: 312.5008 - val mae: 312.5008
Epoch 188/200
29/29 [============== ] - 0s 2ms/step - loss: 315.5713 - ma
e: 315.5713 - val loss: 314.4028 - val mae: 314.4028
Epoch 189/200
e: 319.4303 - val loss: 312.5297 - val mae: 312.5297
Epoch 190/200
29/29 [============== ] - Os 2ms/step - loss: 307.3483 - ma
e: 307.3483 - val loss: 312.9057 - val mae: 312.9057
Epoch 191/200
e: 312.0963 - val loss: 313.6368 - val mae: 313.6368
Epoch 192/200
e: 307.8772 - val_loss: 313.3136 - val_mae: 313.3136
Epoch 193/200
e: 312.9593 - val loss: 312.4965 - val mae: 312.4965
Epoch 194/200
29/29 [=============== ] - 0s 2ms/step - loss: 306.7068 - ma
e: 306.7068 - val loss: 313.8052 - val mae: 313.8052
Epoch 195/200
e: 309.7735 - val loss: 313.4124 - val mae: 313.4124
Epoch 196/200
e: 309.0647 - val loss: 315.7021 - val mae: 315.7021
Epoch 197/200
29/29 [============== ] - Os 2ms/step - loss: 323.6757 - ma
e: 323.6757 - val loss: 312.7341 - val mae: 312.7341
Epoch 198/200
e: 315.4148 - val loss: 313.8771 - val mae: 313.8771
Epoch 199/200
e: 321.5677 - val loss: 313.3227 - val mae: 313.3227
Epoch 200/200
```

```
In [210...
      model = Sequential()
      d rate = 0.8
      reg = 0
      model.add(Dense(256, activation='relu',kernel_regularizer=regularizers.l2(
      Dropout(d_rate)
      model.add(Dense(128,activation='relu', kernel regularizer=regularizers.l2(re
      Dropout(d rate)
      model.add(Dense(128,activation='relu',kernel_regularizer=regularizers.l2(re
      Dropout(d rate)
      model.add(Dense(16,activation='relu',kernel regularizer=regularizers.l2(regularizer)
      model.add(Dense(1, activation='relu'))
      # Compile model
      model.compile(loss='mae', optimizer='adam', metrics=['mae'])
      # Fit the model
      history = model.fit(X_train, y_train, epochs=50, batch_size=100,
                     validation data=(X test, y test),
                    verbose=1)
      # evaluate the model
      scores = model.evaluate(X test, y test, verbose=0)
      print("%s:" % (model.metrics_names[1]))
      Epoch 1/50
      e: 1311.8399 - val_loss: 1315.0153 - val_mae: 1315.0153
      Epoch 2/50
      e: 1258.0613 - val loss: 1082.3937 - val mae: 1082.3937
      e: 887.2685 - val loss: 602.7976 - val_mae: 602.7976
      Epoch 4/50
      e: 551.9157 - val loss: 472.1908 - val mae: 472.1908
      Epoch 5/50
      29/29 [============== ] - Os 2ms/step - loss: 474.1234 - ma
      e: 474.1234 - val loss: 424.9754 - val mae: 424.9754
      Epoch 6/50
      e: 448.5887 - val loss: 399.6656 - val mae: 399.6656
      Epoch 7/50
      29/29 [=============== ] - 0s 2ms/step - loss: 398.7947 - ma
      e: 398.7947 - val loss: 386.5592 - val mae: 386.5592
      Epoch 8/50
      e: 380.4276 - val loss: 382.0818 - val mae: 382.0818
      Epoch 9/50
      e: 384.9023 - val_loss: 371.1081 - val_mae: 371.1081
      Epoch 10/50
      e: 373.5423 - val loss: 365.7001 - val mae: 365.7001
      Epoch 11/50
      e: 377.4002 - val loss: 361.4830 - val mae: 361.4830
      Epoch 12/50
      e: 375.5995 - val_loss: 357.9482 - val_mae: 357.9482
      e: 365.5706 - val loss: 356.0213 - val mae: 356.0213
      Epoch 14/50
```

```
e: 359.5289 - val_loss: 352.7064 - val_mae: 352.7064
Epoch 15/50
29/29 [=============== ] - 0s 2ms/step - loss: 354.9705 - ma
e: 354.9705 - val_loss: 349.8869 - val_mae: 349.8869
Epoch 16/50
29/29 [============== ] - Os 2ms/step - loss: 367.9488 - ma
e: 367.9488 - val_loss: 348.3979 - val_mae: 348.3979
Epoch 17/50
29/29 [=============== ] - 0s 2ms/step - loss: 345.9170 - ma
e: 345.9170 - val_loss: 343.0591 - val_mae: 343.0591
e: 365.3685 - val_loss: 341.3705 - val_mae: 341.3706
Epoch 19/50
e: 354.5482 - val_loss: 340.5789 - val_mae: 340.5789
Epoch 20/50
29/29 [============== ] - 0s 2ms/step - loss: 353.6139 - ma
e: 353.6139 - val loss: 338.6929 - val mae: 338.6929
Epoch 21/50
e: 346.5011 - val loss: 337.4329 - val mae: 337.4329
Epoch 22/50
e: 352.2891 - val_loss: 334.9860 - val_mae: 334.9861
Epoch 23/50
e: 360.6500 - val loss: 334.6175 - val mae: 334.6175
Epoch 24/50
29/29 [============== ] - Os 2ms/step - loss: 346.0458 - ma
e: 346.0458 - val loss: 335.3712 - val mae: 335.3712
Epoch 25/50
e: 349.6846 - val loss: 332.7960 - val mae: 332.7960
Epoch 26/50
e: 344.5929 - val loss: 332.4522 - val mae: 332.4522
Epoch 27/50
e: 333.7929 - val loss: 330.9138 - val mae: 330.9138
Epoch 28/50
29/29 [============== ] - Os 2ms/step - loss: 330.0481 - ma
e: 330.0481 - val loss: 331.6291 - val mae: 331.6290
Epoch 29/50
e: 354.0622 - val loss: 330.1112 - val mae: 330.1112
Epoch 30/50
29/29 [============== ] - Os 2ms/step - loss: 340.2755 - ma
e: 340.2755 - val loss: 327.7443 - val mae: 327.7443
e: 328.6991 - val loss: 325.5136 - val mae: 325.5136
Epoch 32/50
e: 346.1663 - val_loss: 331.1355 - val_mae: 331.1355
Epoch 33/50
e: 323.6632 - val loss: 328.5889 - val mae: 328.5889
Epoch 34/50
29/29 [=============== ] - 0s 2ms/step - loss: 340.1898 - ma
e: 340.1898 - val loss: 325.4608 - val mae: 325.4608
Epoch 35/50
```

```
e: 331.9442 - val loss: 325.4066 - val mae: 325.4066
     Epoch 36/50
     29/29 [=============== ] - 0s 2ms/step - loss: 330.9615 - ma
     e: 330.9615 - val_loss: 322.2797 - val_mae: 322.2797
     Epoch 37/50
     e: 346.8907 - val_loss: 321.6904 - val_mae: 321.6904
     29/29 [============== ] - 0s 2ms/step - loss: 337.3550 - ma
     e: 337.3550 - val loss: 322.7086 - val mae: 322.7086
     Epoch 39/50
     29/29 [============== ] - Os 2ms/step - loss: 331.0273 - ma
     e: 331.0273 - val_loss: 323.8998 - val_mae: 323.8998
     Epoch 40/50
     e: 324.9168 - val loss: 320.7634 - val mae: 320.7634
     Epoch 41/50
     e: 330.1388 - val loss: 319.5902 - val mae: 319.5902
     Epoch 42/50
     e: 327.9298 - val loss: 325.8666 - val mae: 325.8666
     e: 329.0054 - val loss: 319.6252 - val mae: 319.6252
     Epoch 44/50
     e: 332.3099 - val loss: 320.5648 - val mae: 320.5648
     e: 329.9113 - val loss: 319.8901 - val mae: 319.8901
     Epoch 46/50
     29/29 [============== ] - Os 2ms/step - loss: 325.6853 - ma
     e: 325.6853 - val_loss: 322.2478 - val_mae: 322.2478
     Epoch 47/50
     29/29 [============= ] - Os 2ms/step - loss: 339.2200 - ma
     e: 339.2200 - val loss: 322.2568 - val mae: 322.2568
     Epoch 48/50
     e: 328.1887 - val loss: 317.7568 - val mae: 317.7568
     Epoch 49/50
     e: 338.9527 - val loss: 318.4319 - val mae: 318.4319
     Epoch 50/50
     e: 335.9918 - val loss: 317.7234 - val mae: 317.7234
In [198... | model.evaluate(X test, y test)
     ae: 318.9459
Out[198... [318.9458923339844, 318.9458923339844]
In [199... | model.evaluate(X_train, y_train)
     ae: 326.8013
Out[199... [326.8013000488281, 326.8013000488281]
```

Comparision with Regression

```
In [206...
          regr = linear_model.LinearRegression()
          regr.fit(X_train, y_train)
          y predictions = regr.predict(X test)
          print('Mean absolute error: %.2f'
                % mean_absolute_error(y_test, y_predictions))
         Mean absolute error: 597.42
```

In [208... print(y_train.describe())

```
pageviews
count 2881.000000
       1307.810135
mean
       1150.558002
std
min
        104.000000
25%
        554.000000
50%
        936.000000
75%
       1520.000000
       7395.000000
max
```

Conclusion

The mean absolute error with the neural network approach is around 318. The average value for the page views is 1307, this results in an accuracy of around 80% 1307/(1307+318) = 0.804. This result seems acceptable, and is far better than a 'basic' multivariate regression.

The factors wind direction and wind gusts did not seem to be very helpful, and removing them increased the performance.

The model was trained with the factors wind speed, air temperature and location. A further improvement could be the inclusion of the date as an input.

```
In [ ]:
```