In [1]:

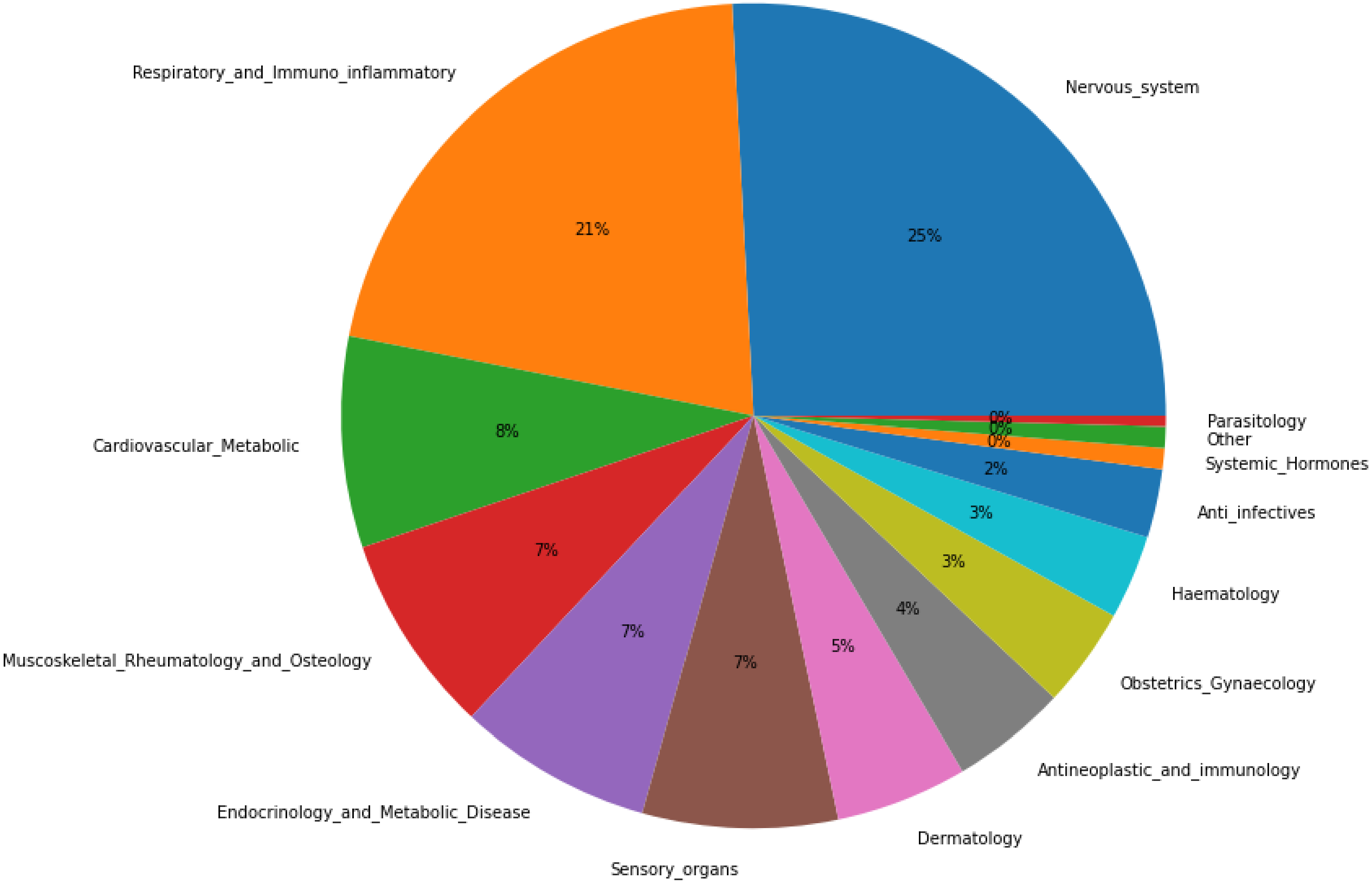
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
plt.rcParams['figure.figsize'] = [10, 10]
```

```
C:\Users\lucia\Anaconda3\lib\site-packages\numpy\_distributor_init.py:32: UserWarning: loaded more than 1 DLL from .l
ibs:
C:\Users\lucia\Anaconda3\lib\site-packages\numpy\.libs\libopenblas.PYQHXLVVQ7VESDPUVUADXEVJOBGHJPAY.gfortran-win_amd6
4.dll
C:\Users\lucia\Anaconda3\lib\site-packages\numpy\.libs\libopenblas.TXA6YQSD3GCQQC22GEQ54J2UDCXDXHWN.gfortran-win_amd6
4.dll
  stacklevel=1)
```

In [2]:

```
#load the data
NumGenericosPorMarca= pd.read_csv('data//gx_num_generics.csv')
FormaDoMedicamento= pd.read_csv('data//gx_package.csv')
CanalDeDistribution= pd.read_csv('data//gx_panel.csv')
AreaTerapeutica= pd.read_csv('data//gx_therapeutic_area.csv') #14 areas
VolumePorMarca= pd.read_csv('data//gx_volume.csv') #atenção que os volumes estão em unidades diferentes por marca
```

```
In [51]: plt.pie(AreaTerapeutica.therapeutic_area.value_counts(), labels= AreaTerapeutica.therapeutic_area.unique(), autopct='%d%%')
         plt.axis('equal')
         plt.show()
```



```
In [55]: VolumePorMarca.describe() #vai de -137 meses a 23 meses depois da introdução do genérico # tem volumes de zero!!
```

Out[55]:

	Unnamed: 0	volume	month_num
count	94954.000000	9.495400e+04	94954.000000
mean	47477.500000	7.288071e+07	-30.983287
std	27411.003068	1.041108e+09	34.773676
min	1.000000	0.000000e+00	-137.000000
25%	23739.250000	9.025609e+05	-54.000000
50%	47477.500000	4.861060e+06	-26.000000
75%	71215.750000	2.078613e+07	-3.000000
max	94954.000000	6.041151e+10	23.000000

```
In [51]: len(VolumePorMarca)

Out[51]: 94954
```

```
In [69]: len(VolumePorMarca.brand.unique()) #484 produtos

Out[69]: 484
```

```
In [70]: len(VolumePorMarca.country.unique()) # 16 países

Out[70]: 16
```

In [50]:

```
FormaDoMedicamento= FormaDoMedicamento.iloc[:,1:]
FormaDoMedicamento.head()
```

Out[50]:

	country	brand	presentation
0	country_1	brand_3	PILL
1	country_1	brand_4	PILL
2	country_1	brand_10	PILL
3	country_1	brand_14	PILL
4	country_1	brand_18	CREAM

In [54]:

```
len(FormaDoMedicamento[(FormaDoMedicamento.presentation =='PILL')])
```

Out[54]:

```
677
```

In [80]:

```
FormaDoMedicamento[(FormaDoMedicamento.country=='country_9')].head(20) # num pais_brand pair a forma do medicamento é
 unica
```

Out[80]:

	country	brand	presentation
607	country_9	brand_167	PILL
608	country_9	brand_187	PILL
609	country_9	brand_477	INJECTION

In [58]:

```
FormaDoMedicamento.presentation.value_counts() #7 tipos diferentes
```

Out[58]:

```
PILL          677
OTHER         233
INJECTION      91
EYE_DROP       34
CREAM          29
PATCH          13
INHALER         1
Name: presentation, dtype: int64
```

In [81]:

```
#CanalDeDistribution.head()
CanalDeDistribution=CanalDeDistribution.iloc[:,1:]
CanalDeDistribution.head()
```

Out[81]:

	country	brand	channel	channel_rate
0	country_1	brand_3	B	1.189704
1	country_1	brand_3	D	98.810296
2	country_1	brand_4	B	0.090229
3	country_1	brand_4	D	99.909771
4	country_1	brand_10	B	1.015697

In [106]:

```
Dummies_DistrChan=pd.get_dummies(CanalDeDistribution['channel'])
Dummies_DistrChan2 = Dummies_DistrChan.join(CanalDeDistribution[['country', 'brand', 'channel_rate']])
#Dummies_DistrChan2.head()
```

In [92]:

```
Dummies_DistrChan2['A']=Dummies_DistrChan2['A']*Dummies_DistrChan2['channel_rate']/100
Dummies_DistrChan2['B']=Dummies_DistrChan2['B']*Dummies_DistrChan2['channel_rate']/100
Dummies_DistrChan2['C']=Dummies_DistrChan2['C']*Dummies_DistrChan2['channel_rate']/100
Dummies_DistrChan2['D']=Dummies_DistrChan2['D']*Dummies_DistrChan2['channel_rate']/100
Dummies_DistrChan2.head()
```

Out[92]:

	A	B	C	D	country	brand	channel_rate
0	0.0	0.011897	0.0	0.000000	country_1	brand_3	1.189704
1	0.0	0.000000	0.0	0.988103	country_1	brand_3	98.810296
2	0.0	0.000902	0.0	0.000000	country_1	brand_4	0.090229
3	0.0	0.000000	0.0	0.999098	country_1	brand_4	99.909771
4	0.0	0.010157	0.0	0.000000	country_1	brand_10	1.015697

In [111]:

```
b=Dummies_DistrChan2.groupby(['country', 'brand']).sum()
b.head()
```

Out[111]:

		A	B	C	D	channel_rate
country	brand					
country_1	brand_10	0.0	0.010157	0.0	0.989843	100.0
	brand_102	0.0	0.001098	0.0	0.998902	100.0
	brand_115	0.0	0.019927	0.0	0.980073	100.0
	brand_117	0.0	0.021485	0.0	0.978515	100.0
	brand_119	0.0	0.006055	0.0	0.993945	100.0

In [121]:

```
Dummies_DistrChan2=Dummies_DistrChan2.drop(columns = ['A', 'B', 'C', 'D', 'channel_rate'])

Dummies_DistrChan2.head()
```

Out[121]:

	country	brand
0	country_1	brand_3
1	country_1	brand_3
2	country_1	brand_4
3	country_1	brand_4
4	country_1	brand_10

In [134]:

```
Dummies_DistrChan2=Dummies_DistrChan2.drop_duplicates()
len(Dummies_DistrChan2)
```

Out[134]:

```
1078
```

In [144]:

```
Dummies_DistrChan4=Dummies_DistrChan2.merge(b, on=['country', 'brand'])
Dummies_DistrChan4=Dummies_DistrChan4.drop(columns=['A_sum', 'B_sum'])
Dummies_DistrChan4.head()
```

Out[144]:

	country	brand	A	B	C	D
0	country_1	brand_3	0.0	0.011897	0.0	0.988103
1	country_1	brand_4	0.0	0.000902	0.0	0.999098
2	country_1	brand_10	0.0	0.010157	0.0	0.989843
3	country_1	brand_14	0.0	0.011184	0.0	0.988816
4	country_1	brand_18	0.0	0.011187	0.0	0.988813

In [60]:

```
NumGenericosPorMarca.describe() #min 1 max 50
```

Out[60]:

	Unnamed: 0	num_generics
count	1078.000000	1078.000000
mean	539.500000	4.790353
std	311.336099	6.718827
min	1.000000	1.000000
25%	270.250000	1.000000
50%	539.500000	1.000000
75%	808.750000	6.000000
max	1078.000000	50.000000

In [150]:

```
NumGenericosPorMarca=NumGenericosPorMarca.iloc[:,1:]
NumGenericosPorMarca.head()
```

Out[150]:

	country	brand	num_generics
0	country_1	brand_3	3
1	country_1	brand_4	1
2	country_1	brand_10	6
3	country_1	brand_14	1
4	country_1	brand_18	1

In [67]:

```
UmGenerico=NumGenericosPorMarca[NumGenericosPorMarca.num_generics==1]
UmGenerico.brand.count()
```

Out[67]:

```
592
```

juntar os datasets

In [4]:

```
CanalDeDistribution=CanalDeDistribution.iloc[:,1:]
#CanalDeDistribution.head()
```

In [146]:

```
Merge2=FormaDoMedicamento.merge(Dummies_DistrChan4, on=['country','brand'])
#Merge2.head()
len(Merge2)
```

Out[146]:

```
1078
```

In [147]:

```
VolumePorMarca=VolumePorMarca.iloc[:,1:]
VolumePorMarca.head()
```

Out[147]:

	country	brand	volume	month_num	month_name
0	country_1	brand_3	18509088.6	-88	Jul
1	country_1	brand_3	19697508.0	-87	Aug
2	country_1	brand_3	18315721.8	-86	Sep
3	country_1	brand_3	19831199.4	-85	Oct
4	country_1	brand_3	18593281.8	-84	Nov

In [149]:

```
Merge3=Merge2.merge(VolumePorMarca, on=['country','brand'])
#len(Merge3) #94954
Merge3.head()
```

Out[149]:

	country	brand	presentation	A	B	C	D	volume	month_num	month_name
0	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18509088.6	-88	Jul
1	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19697508.0	-87	Aug
2	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18315721.8	-86	Sep
3	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19831199.4	-85	Oct
4	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18593281.8	-84	Nov

In [154]:

```
Merge4=Merge3.merge(NumGenericosPorMarca, on=['country','brand'] )
Merge4.head()
#len(Merge4) #94954
```

Out[154]:

	country	brand	presentation	A	B	C	D	volume	month_num	month_name	num_generics
0	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18509088.6	-88	Jul	3
1	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19697508.0	-87	Aug	3
2	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18315721.8	-86	Sep	3
3	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19831199.4	-85	Oct	3
4	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18593281.8	-84	Nov	3

In [155]:

```
AreaTerapeutica=AreaTerapeutica.iloc[:,1:]
AreaTerapeutica.head()
```

Out[155]:

	brand	therapeutic_area
0	brand_1	Nervous_system
1	brand_2	Respiratory_and_Immuno_inflammatory
2	brand_3	Cardiovascular_Metabolic
3	brand_4	Cardiovascular_Metabolic
4	brand_5	Cardiovascular_Metabolic

In [156]:

```
FinalData=Merge4.merge(AreaTerapeutica, on='brand')
len(FinalData)
```

Out[156]:

```
94954
```

In [157]:

```python
FinalData.head()
```

Out[157]:

	country	brand	presentation	A	B	C	D	volume	month_num	month_name	num_generics	therapeutic_area
0	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18509088.6	-88	Jul	3	Cardiovascular_Metabolic
1	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19697508.0	-87	Aug	3	Cardiovascular_Metabolic
2	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18315721.8	-86	Sep	3	Cardiovascular_Metabolic
3	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19831199.4	-85	Oct	3	Cardiovascular_Metabolic
4	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18593281.8	-84	Nov	3	Cardiovascular_Metabolic

Make it supervised

In [162]:

```python
FinalData['VolTplus1']= FinalData.groupby(['country', 'brand']).volume.shift(-1)
FinalData.head()
#len(FinalData) #94954
```

Out[162]:

	country	brand	presentation	A	B	C	D	volume	month_num	month_name	num_generics	therapeutic_area	
0	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18509088.6	-88	Jul	3	Cardiovascular_Metabolic	1
1	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19697508.0	-87	Aug	3	Cardiovascular_Metabolic	1
2	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18315721.8	-86	Sep	3	Cardiovascular_Metabolic	1
3	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19831199.4	-85	Oct	3	Cardiovascular_Metabolic	1
4	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18593281.8	-84	Nov	3	Cardiovascular_Metabolic	1

```
In [166]: FinalData.isnull().sum()
```

```
Out[166]: country                 0
          brand                   0
          presentation            0
          A                       0
          B                       0
          C                       0
          D                       0
          volume                  0
          month_num               0
          month_name              0
          num_generics            0
          therapeutic_area        0
          VolTplus1            1078
          log_vol                 0
          log_volTplus1        1078
          dtype: int64
```

calculate log version of the units sold

In [165]:

```
FinalData['log_vol']=np.log10(FinalData.volume)
FinalData['log_volTplus1']=np.log10(FinalData.VolTplus1)
FinalData.head()
```

```
C:\Users\lucia\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: RuntimeWarning: divide by zero encountered in log
10
  """Entry point for launching an IPython kernel.
C:\Users\lucia\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: RuntimeWarning: divide by zero encountered in log
10
```

Out[165]:

	country	brand	presentation	A	B	C	D	volume	month_num	month_name	num_generics	therapeutic_area	
0	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18509088.6	-88	Jul	3	Cardiovascular_Metabolic	1
1	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19697508.0	-87	Aug	3	Cardiovascular_Metabolic	1
2	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18315721.8	-86	Sep	3	Cardiovascular_Metabolic	1
3	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19831199.4	-85	Oct	3	Cardiovascular_Metabolic	1
4	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18593281.8	-84	Nov	3	Cardiovascular_Metabolic	1

In [170]:

```
FinalDummies=['country', 'brand', 'presentation', 'month_name', 'therapeutic_area']
DummyFeatures=pd.get_dummies(FinalData[FinalDummies])
DummyFeatures.head() #553 colunas
```

Out[170]:

	country_country_1	country_country_10	country_country_11	country_country_12	country_country_13	country_country_14	country_country_15
0	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0
2	1	0	0	0	0	0	0
3	1	0	0	0	0	0	0
4	1	0	0	0	0	0	0

5 rows × 533 columns

In [177]:
```
Final_Superv=pd.merge(FinalData, DummyFeatures, left_on=FinalData.index, right_on=DummyFeatures.index)
Final_Superv.head() #549 colunas
#len(Final_Superv) #94954
```

Out[177]:

	key_0	country	brand	presentation	A	B	C	D	volume	month_num	...	therapeutic_area_Endocrinology_and_Metaboli
0	0	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18509088.6	-88	...	
1	1	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19697508.0	-87	...	
2	2	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18315721.8	-86	...	
3	3	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	19831199.4	-85	...	
4	4	country_1	brand_3	PILL	0.0	0.011897	0.0	0.988103	18593281.8	-84	...	

5 rows × 549 columns

In [188]:
```
list(Final_Superv.columns)[541:]
```

Out[188]:
```
['therapeutic_area_Muscoskeletal_Rheumatology_and_Osteology',
 'therapeutic_area_Nervous_system',
 'therapeutic_area_Obstetrics_Gynaecology',
 'therapeutic_area_Other',
 'therapeutic_area_Parasitology',
 'therapeutic_area_Respiratory_and_Immuno_inflammatory',
 'therapeutic_area_Sensory_organs',
 'therapeutic_area_Systemic_Hormones']
```

In [190]:
```
RelevantFeatures=Final_Superv.drop(columns=['key_0', 'country', 'brand', 'presentation', 'volume', 'month_name', 'ther
apeutic_area',
                                            'VolTplus1'])
RelevantFeatures.head()
```

Out[190]:
```
     A         B    C         D  month_num  num_generics   log_vol  log_volTplus1  country_country_1  country_country_10  ...  therapeutic_area_
0  0.0  0.011897  0.0  0.988103        -88             3  7.267385       7.294411                  1                   0  ...
1  0.0  0.011897  0.0  0.988103        -87             3  7.294411       7.262824                  1                   0  ...
2  0.0  0.011897  0.0  0.988103        -86             3  7.262824       7.297349                  1                   0  ...
3  0.0  0.011897  0.0  0.988103        -85             3  7.297349       7.269356                  1                   0  ...
4  0.0  0.011897  0.0  0.988103        -84             3  7.269356       7.283714                  1                   0  ...

5 rows × 541 columns
```

In [191]:
```
#separate test # remove 15 brands out of the 484
EvalSet= RelevantFeatures.loc[(RelevantFeatures.brand_brand_451 == 1) & (RelevantFeatures.month_num > -1)]
```

In [192]:
```
EvalSeta = RelevantFeatures.loc[(RelevantFeatures.brand_brand_131 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetb = RelevantFeatures.loc[(RelevantFeatures.brand_brand_227 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetc = RelevantFeatures.loc[(RelevantFeatures.brand_brand_310 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetd = RelevantFeatures.loc[(RelevantFeatures.brand_brand_101 == 1) & (RelevantFeatures.month_num > -1)]
EvalSete = RelevantFeatures.loc[(RelevantFeatures.brand_brand_233 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetf = RelevantFeatures.loc[(RelevantFeatures.brand_brand_215 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetg = RelevantFeatures.loc[(RelevantFeatures.brand_brand_53 == 1) & (RelevantFeatures.month_num > -1)]
EvalSeth = RelevantFeatures.loc[(RelevantFeatures.brand_brand_337 == 1) & (RelevantFeatures.month_num > -1)]
EvalSeti = RelevantFeatures.loc[(RelevantFeatures.brand_brand_447 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetj = RelevantFeatures.loc[(RelevantFeatures.brand_brand_478 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetk = RelevantFeatures.loc[(RelevantFeatures.brand_brand_54 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetl = RelevantFeatures.loc[(RelevantFeatures.brand_brand_161 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetm = RelevantFeatures.loc[(RelevantFeatures.brand_brand_403 == 1) & (RelevantFeatures.month_num > -1)]
EvalSetn = RelevantFeatures.loc[(RelevantFeatures.brand_brand_443 == 1) & (RelevantFeatures.month_num > -1)]
EvalSeto = RelevantFeatures.loc[(RelevantFeatures.brand_brand_385 == 1) & (RelevantFeatures.month_num > -1)]
```

In [204]:

```
frames2=[EvalSet, EvalSeta,EvalSetb,EvalSetc,EvalSetd,EvalSete,EvalSetf,EvalSetg,EvalSeth,EvalSeti,EvalSetj,EvalSetk,EvalSetl,EvalSetm,EvalSetn, EvalSeto]
Eval_Set=pd.concat(frames2)
```

In [194]:

```
#drop evaluation instances
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_451 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_131 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_227 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_310 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_101 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_233 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_215 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_53 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_337 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_447 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_478 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_54 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_161 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_403 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_443 == 1) & (RelevantFeatures.month_num > -1))].index)
RelevantFeatures= RelevantFeatures.drop(RelevantFeatures[((RelevantFeatures.brand_brand_385 == 1) & (RelevantFeatures.month_num > -1))].index)
```

In [205]:

```
Eval_Set.head()
```

Out[205]:

	A	B	C	D	month_num	num_generics	log_vol	log_volTplus1	country_country_1	country_country_10	...	therapeutic_
40714	0.0	0.033377	0.0	0.966623	0	2	7.350803	7.159221	1	0	...	
40715	0.0	0.033377	0.0	0.966623	1	2	7.159221	7.112165	1	0	...	
40716	0.0	0.033377	0.0	0.966623	2	2	7.112165	7.136766	1	0	...	
40717	0.0	0.033377	0.0	0.966623	3	2	7.136766	7.137970	1	0	...	
40718	0.0	0.033377	0.0	0.966623	4	2	7.137970	7.162980	1	0	...	

5 rows × 541 columns

Random Forest Regressor

In [281]:

```
#remove NaN RF does not handle it...
RelevantFeatures=RelevantFeatures.replace('-inf', np.nan)
RelevantFeatures2=RelevantFeatures.dropna()
Eval_Set2=Eval_Set.dropna()
```

In [246]:

```
train_X= RelevantFeatures2.drop('log_volTplus1', axis=1)
train_y= RelevantFeatures2.log_volTplus1
teste_X=Eval_Set2.drop('log_volTplus1', axis=1)
teste_y=Eval_Set2.log_volTplus1
```

In [270]:

```
len(train_X)
```

Out[270]:

```
92604
```

In [250]:

```python
from sklearn.ensemble import RandomForestRegressor
modelRF=RandomForestRegressor(random_state=0, verbose=2, n_jobs=-1)
modelRF.fit(train_X, train_y)
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 3 of 100building tree 1 of 100building tree 2 of 100building tree 4 of 100building tree 5 of 100
building tree 6 of 100



building tree 8 of 100building tree 7 of 100


building tree 9 of 100
building tree 10 of 100
building tree 11 of 100
building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100
building tree 36 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   23.6s
```

```
building tree 37 of 100
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100
building tree 50 of 100
building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
building tree 74 of 100
building tree 75 of 100
building tree 76 of 100
building tree 77 of 100
building tree 78 of 100
building tree 79 of 100
building tree 80 of 100
building tree 81 of 100
building tree 82 of 100
building tree 83 of 100
building tree 84 of 100
building tree 85 of 100
building tree 86 of 100
building tree 87 of 100
building tree 88 of 100
building tree 89 of 100
building tree 90 of 100
building tree 91 of 100
building tree 92 of 100
building tree 93 of 100
building tree 94 of 100
building tree 95 of 100
building tree 96 of 100
building tree 97 of 100
building tree 98 of 100
building tree 99 of 100
building tree 100 of 100

[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:  1.1min finished
```

Out[250]:

```
RandomForestRegressor(n_jobs=-1, random_state=0, verbose=2)
```

In [251]:

```
from sklearn.metrics import mean_squared_error
pred_train=modelRF.predict(train_X)
mean_squared_error(train_y, pred_train)
```

```
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    1.1s finished
```

Out[251]:

```
0.000665892796098858
```

In [252]:

```
pred_test=modelRF.predict(teste_X)
mean_squared_error(teste_y, pred_test)

#seems promising
```

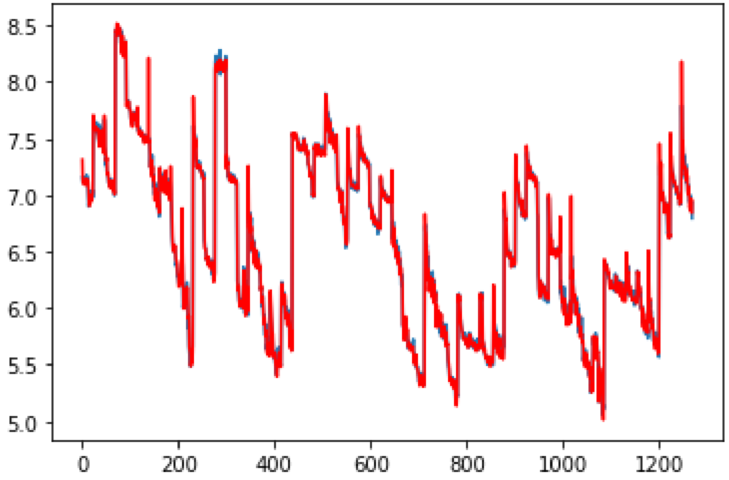
```
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.1s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.2s finished
```

Out[252]:

```
0.006061211175460223
```

In [255]:

```
#import matplotlib.pyplot as plt
plt.plot(teste_y.values)
plt.plot(pred_test, c='r')
plt.show()
```



In [264]:

```
#importances = modelRF.feature_importances_
#indices = np.argsort(importances)[::-1]
#plt.bar(range(teste_X.shape[1]), importances[indices],
        #color="r", align="center")
#importances[indices]
#plt.xticks(range(teste_X.shape[1]), indices)
#plt.show()
#indices #6= log_vol
```

In [267]:

```
#agora tenho de esconder a informação verdadeira do log_vol e substituir pelo valor previsto
Eval_month0=Eval_Set2.loc[Eval_Set2.month_num == 0]
Eval_month1=Eval_Set2.loc[Eval_Set2.month_num == 1]
Eval_month2=Eval_Set2.loc[Eval_Set2.month_num == 2]
Eval_month3=Eval_Set2.loc[Eval_Set2.month_num == 3]
Eval_month4=Eval_Set2.loc[Eval_Set2.month_num == 4]
Eval_month5=Eval_Set2.loc[Eval_Set2.month_num == 5]
Eval_month6=Eval_Set2.loc[Eval_Set2.month_num == 6]
Eval_month7=Eval_Set2.loc[Eval_Set2.month_num == 7]
Eval_month8=Eval_Set2.loc[Eval_Set2.month_num == 8]
Eval_month9=Eval_Set2.loc[Eval_Set2.month_num == 9]
Eval_month10=Eval_Set2.loc[Eval_Set2.month_num == 10]
Eval_month11=Eval_Set2.loc[Eval_Set2.month_num == 11]
Eval_month12=Eval_Set2.loc[Eval_Set2.month_num == 12]
Eval_month13=Eval_Set2.loc[Eval_Set2.month_num == 13]
Eval_month14=Eval_Set2.loc[Eval_Set2.month_num == 14]
Eval_month15=Eval_Set2.loc[Eval_Set2.month_num == 15]
Eval_month16=Eval_Set2.loc[Eval_Set2.month_num == 16]
Eval_month17=Eval_Set2.loc[Eval_Set2.month_num == 17]
Eval_month18=Eval_Set2.loc[Eval_Set2.month_num == 18]
Eval_month19=Eval_Set2.loc[Eval_Set2.month_num == 19]
Eval_month20=Eval_Set2.loc[Eval_Set2.month_num == 20]
Eval_month21=Eval_Set2.loc[Eval_Set2.month_num == 21]
Eval_month22=Eval_Set2.loc[Eval_Set2.month_num == 22]
Eval_month23=Eval_Set2.loc[Eval_Set2.month_num == 23]
```

In [268]:

```
teste_X_0=Eval_month0.drop('log_volTplus1', axis=1)
teste_y_0=Eval_month0.log_volTplus1
pred_test_0=modelRF.predict(teste_X_0)
error_0=mean_squared_error(teste_y_0, pred_test_0)
error_0
```

```
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.0s finished
```

Out[268]:

```
0.05225525267710426
```

In [286]:

```
teste_X_0.log_vol.replace(teste_X_0.log_vol.values, pred_test_0, inplace=True)
teste_X_0['log_volTplus1']= Eval_month0.log_volTplus1
#teste_X_0.head()
RelevantFeatures_month_1= RelevantFeatures2.append(teste_X_0, ignore_index=True, sort=False)
RelevantFeatures_month_1.head()
#len(RelevantFeatures_month_1) #56+ 92604= 92660

train_X_1= RelevantFeatures_month_1.drop('log_volTplus1', axis=1)
train_y_1= RelevantFeatures_month_1.log_volTplus1
teste_X_1=Eval_month1.drop('log_volTplus1', axis=1)
teste_y_1=Eval_month1.log_volTplus1

modelRF.fit(train_X_1, train_y_1)
pred_test_1=modelRF.predict(teste_X_1)
error_1=mean_squared_error(teste_y_1, pred_test_1)
error_1
```

```
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.1s finished
```

Out[286]:

```
0.011459254654542268
```

In [287]:

```
teste_X_1.log_vol.replace(teste_X_1.log_vol.values, pred_test_1, inplace=True)
teste_X_1['log_volTplus1']= Eval_month1.log_volTplus1

RelevantFeatures_month_2= RelevantFeatures_month_1.append(teste_X_1, ignore_index=True, sort=False)


train_X_2= RelevantFeatures_month_2.drop('log_volTplus1', axis=1)
train_y_2= RelevantFeatures_month_2.log_volTplus1
teste_X_2=Eval_month2.drop('log_volTplus1', axis=1)
teste_y_2=Eval_month2.log_volTplus1

modelRF.fit(train_X_2, train_y_2)
pred_test_2=modelRF.predict(teste_X_2)
error_2=mean_squared_error(teste_y_2, pred_test_2)
error_2
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 1 of 100
building tree 2 of 100
building tree 3 of 100
building tree 4 of 100
building tree 5 of 100building tree 6 of 100

building tree 7 of 100
building tree 8 of 100
building tree 9 of 100
building tree 10 of 100
building tree 11 of 100
building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   17.6s
```

```
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100
building tree 36 of 100
building tree 37 of 100
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100building tree 50 of 100

building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
building tree 74 of 100
building tree 75 of 100
building tree 76 of 100
building tree 77 of 100
building tree 78 of 100
building tree 79 of 100
building tree 80 of 100
building tree 81 of 100
building tree 82 of 100
building tree 83 of 100
building tree 84 of 100
building tree 85 of 100
building tree 86 of 100
building tree 87 of 100
building tree 88 of 100
building tree 89 of 100
building tree 90 of 100
building tree 91 of 100
building tree 92 of 100
building tree 93 of 100
building tree 94 of 100
building tree 95 of 100
building tree 96 of 100
building tree 97 of 100
building tree 98 of 100
building tree 99 of 100
building tree 100 of 100

[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:   59.0s finished
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.0s finished
```

```
Out[287]: 0.005987474679851099
```

In [289]:

```
teste_X_2.log_vol.replace(teste_X_2.log_vol.values, pred_test_2, inplace=True)
teste_X_2['log_volTplus1']= Eval_month2.log_volTplus1

RelevantFeatures_month_3= RelevantFeatures_month_2.append(teste_X_2, ignore_index=True, sort=False)


train_X_3= RelevantFeatures_month_3.drop('log_volTplus1', axis=1)
train_y_3= RelevantFeatures_month_3.log_volTplus1
teste_X_3=Eval_month3.drop('log_volTplus1', axis=1)
teste_y_3=Eval_month3.log_volTplus1

modelRF.fit(train_X_3, train_y_3)
pred_test_3=modelRF.predict(teste_X_3)
error_3=mean_squared_error(teste_y_3, pred_test_3)
error_3
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 2 of 100building tree 3 of 100
building tree 4 of 100
building tree 1 of 100
building tree 5 of 100
building tree 6 of 100building tree 8 of 100

building tree 7 of 100

building tree 9 of 100building tree 10 of 100building tree 11 of 100
building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100

building tree 16 of 100

building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   21.4s
```

```
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100
building tree 36 of 100
building tree 37 of 100
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100
building tree 50 of 100
building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
```

```
          building tree 74 of 100
          building tree 75 of 100
          building tree 76 of 100
          building tree 77 of 100
          building tree 78 of 100
          building tree 79 of 100
          building tree 80 of 100
          building tree 81 of 100
          building tree 82 of 100
          building tree 83 of 100
          building tree 84 of 100
          building tree 85 of 100
          building tree 86 of 100
          building tree 87 of 100
          building tree 88 of 100
          building tree 89 of 100
          building tree 90 of 100
          building tree 91 of 100
          building tree 92 of 100
          building tree 93 of 100
          building tree 94 of 100
          building tree 95 of 100
          building tree 96 of 100
          building tree 97 of 100
          building tree 98 of 100
          building tree 99 of 100
          building tree 100 of 100

          [Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:  1.1min finished
          [Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
          [Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.0s
          [Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.1s finished

Out[289]: 0.003831693651965576
```

In [294]:

```
teste_X_3.log_vol.replace(teste_X_3.log_vol.values, pred_test_3, inplace=True)
teste_X_3['log_volTplus1']= Eval_month3.log_volTplus1

RelevantFeatures_month_4= RelevantFeatures_month_3.append(teste_X_3, ignore_index=True, sort=False)



train_X_4= RelevantFeatures_month_4.drop('log_volTplus1', axis=1)
train_y_4= RelevantFeatures_month_4.log_volTplus1
teste_X_4=Eval_month4.drop('log_volTplus1', axis=1)
teste_y_4=Eval_month4.log_volTplus1

modelRF.fit(train_X_4, train_y_4)
pred_test_4=modelRF.predict(teste_X_4)
error_4=mean_squared_error(teste_y_4, pred_test_4)
error_4
```

In [295]:

```
teste_X_4.log_vol.replace(teste_X_4.log_vol.values, pred_test_4, inplace=True)
teste_X_4['log_volTplus1']= Eval_month4.log_volTplus1

RelevantFeatures_month_5= RelevantFeatures_month_4.append(teste_X_4, ignore_index=True, sort=False)



train_X_5= RelevantFeatures_month_5.drop('log_volTplus1', axis=1)
train_y_5= RelevantFeatures_month_5.log_volTplus1
teste_X_5=Eval_month5.drop('log_volTplus1', axis=1)
teste_y_5=Eval_month5.log_volTplus1

modelRF.fit(train_X_5, train_y_5)
pred_test_5=modelRF.predict(teste_X_5)
error_5=mean_squared_error(teste_y_5, pred_test_5)
error_5
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 2 of 100building tree 1 of 100building tree 3 of 100building tree 4 of 100building tree 5 of 100



building tree 6 of 100building tree 7 of 100

building tree 8 of 100

building tree 9 of 100
building tree 10 of 100
building tree 11 of 100
building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   18.3s
```

```
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100
building tree 36 of 100
building tree 37 of 100
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100
building tree 50 of 100
building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
```

```
          building tree 74 of 100
          building tree 75 of 100
          building tree 76 of 100
          building tree 77 of 100
          building tree 78 of 100
          building tree 79 of 100
          building tree 80 of 100
          building tree 81 of 100
          building tree 82 of 100
          building tree 83 of 100
          building tree 84 of 100
          building tree 85 of 100
          building tree 86 of 100
          building tree 87 of 100
          building tree 88 of 100
          building tree 89 of 100
          building tree 90 of 100
          building tree 91 of 100
          building tree 92 of 100
          building tree 93 of 100
          building tree 94 of 100
          building tree 95 of 100
          building tree 96 of 100
          building tree 97 of 100
          building tree 98 of 100
          building tree 99 of 100
          building tree 100 of 100

          [Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:  1.0min finished
          [Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
          [Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.1s
          [Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.2s finished

Out[295]: 0.0038297763555058526
```

In [297]:

```
teste_X_5.log_vol.replace(teste_X_5.log_vol.values, pred_test_5, inplace=True)
teste_X_5['log_volTplus1']= Eval_month5.log_volTplus1

RelevantFeatures_month_6= RelevantFeatures_month_5.append(teste_X_5, ignore_index=True, sort=False)


train_X_6= RelevantFeatures_month_6.drop('log_volTplus1', axis=1)
train_y_6= RelevantFeatures_month_6.log_volTplus1
teste_X_6=Eval_month6.drop('log_volTplus1', axis=1)
teste_y_6=Eval_month6.log_volTplus1

modelRF.fit(train_X_6, train_y_6)
pred_test_6=modelRF.predict(teste_X_6)
error_6=mean_squared_error(teste_y_5, pred_test_6)
error_6
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 1 of 100
building tree 2 of 100
building tree 3 of 100building tree 4 of 100building tree 5 of 100

building tree 6 of 100
building tree 7 of 100building tree 8 of 100


building tree 9 of 100
building tree 10 of 100building tree 11 of 100

building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   18.5s
```

```
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100
building tree 36 of 100
building tree 37 of 100
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100
building tree 50 of 100
building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
building tree 74 of 100
building tree 75 of 100
building tree 76 of 100
building tree 77 of 100
building tree 78 of 100
building tree 79 of 100
building tree 80 of 100
building tree 81 of 100
building tree 82 of 100
building tree 83 of 100
building tree 84 of 100
building tree 85 of 100
building tree 86 of 100
building tree 87 of 100
building tree 88 of 100
building tree 89 of 100
building tree 90 of 100
building tree 91 of 100
building tree 92 of 100
building tree 93 of 100
building tree 94 of 100
building tree 95 of 100
building tree 96 of 100
building tree 97 of 100
building tree 98 of 100
building tree 99 of 100
building tree 100 of 100

[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:  1.0min finished
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.0s finished
```

```
Out[297]: 0.0016647004566998846
```

In [298]:

```
teste_X_6.log_vol.replace(teste_X_6.log_vol.values, pred_test_6, inplace=True)
teste_X_6['log_volTplus1']= Eval_month6.log_volTplus1

RelevantFeatures_month_7= RelevantFeatures_month_6.append(teste_X_6, ignore_index=True, sort=False)


train_X_7= RelevantFeatures_month_7.drop('log_volTplus1', axis=1)
train_y_7= RelevantFeatures_month_7.log_volTplus1
teste_X_7=Eval_month7.drop('log_volTplus1', axis=1)
teste_y_7=Eval_month7.log_volTplus1

modelRF.fit(train_X_7, train_y_7)
pred_test_7=modelRF.predict(teste_X_7)
error_7=mean_squared_error(teste_y_7, pred_test_7)
error_7
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 1 of 100building tree 2 of 100building tree 3 of 100
building tree 4 of 100
building tree 5 of 100

building tree 7 of 100
building tree 6 of 100building tree 8 of 100


building tree 9 of 100
building tree 10 of 100
building tree 11 of 100
building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   18.9s
```

```
building tree 36 of 100
building tree 37 of 100
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100
building tree 50 of 100
building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
building tree 74 of 100
building tree 75 of 100
building tree 76 of 100
building tree 77 of 100
building tree 78 of 100
building tree 79 of 100
building tree 80 of 100
building tree 81 of 100
building tree 82 of 100
building tree 83 of 100
building tree 84 of 100
building tree 85 of 100
building tree 86 of 100
building tree 87 of 100
building tree 88 of 100
building tree 89 of 100
building tree 90 of 100
building tree 91 of 100
building tree 92 of 100
building tree 93 of 100
building tree 94 of 100
building tree 95 of 100
building tree 96 of 100
building tree 97 of 100
building tree 98 of 100
building tree 99 of 100
building tree 100 of 100

[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:  1.1min finished
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.5s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.6s finished
```

```
Out[298]: 0.002410150652359036
```

In [299]:

```
teste_X_7.log_vol.replace(teste_X_7.log_vol.values, pred_test_7, inplace=True)
teste_X_7['log_volTplus1']= Eval_month7.log_volTplus1

RelevantFeatures_month_8= RelevantFeatures_month_7.append(teste_X_7, ignore_index=True, sort=False)



train_X_8= RelevantFeatures_month_8.drop('log_volTplus1', axis=1)
train_y_8= RelevantFeatures_month_8.log_volTplus1
teste_X_8=Eval_month8.drop('log_volTplus1', axis=1)
teste_y_8=Eval_month8.log_volTplus1

modelRF.fit(train_X_8, train_y_8)
pred_test_8=modelRF.predict(teste_X_8)
error_8=mean_squared_error(teste_y_8, pred_test_8)
error_8
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 1 of 100building tree 2 of 100
building tree 3 of 100building tree 4 of 100
building tree 5 of 100


building tree 6 of 100
building tree 7 of 100
building tree 8 of 100
building tree 9 of 100
building tree 10 of 100
building tree 11 of 100
building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100
building tree 36 of 100
building tree 37 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   19.9s
```

```
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100
building tree 50 of 100
building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
building tree 74 of 100
building tree 75 of 100
building tree 76 of 100
building tree 77 of 100
building tree 78 of 100
building tree 79 of 100
building tree 80 of 100
building tree 81 of 100
building tree 82 of 100
building tree 83 of 100
building tree 84 of 100
building tree 85 of 100
building tree 86 of 100
building tree 87 of 100
building tree 88 of 100
building tree 89 of 100
building tree 90 of 100
building tree 91 of 100
building tree 92 of 100
building tree 93 of 100
building tree 94 of 100
building tree 95 of 100
building tree 96 of 100
building tree 97 of 100
building tree 98 of 100
building tree 99 of 100
building tree 100 of 100

[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:  1.1min finished
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.2s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.2s finished
```

```
Out[299]: 0.002331916586338828
```

In [300]:

```
teste_X_8.log_vol.replace(teste_X_8.log_vol.values, pred_test_8, inplace=True)
teste_X_8['log_volTplus1']= Eval_month8.log_volTplus1

RelevantFeatures_month_9= RelevantFeatures_month_8.append(teste_X_8, ignore_index=True, sort=False)


train_X_9= RelevantFeatures_month_9.drop('log_volTplus1', axis=1)
train_y_9= RelevantFeatures_month_9.log_volTplus1
teste_X_9=Eval_month9.drop('log_volTplus1', axis=1)
teste_y_9=Eval_month9.log_volTplus1

modelRF.fit(train_X_9, train_y_9)
pred_test_9=modelRF.predict(teste_X_9)
error_9=mean_squared_error(teste_y_9, pred_test_9)
error_9
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 2 of 100building tree 3 of 100building tree 1 of 100building tree 4 of 100

building tree 5 of 100

building tree 6 of 100building tree 7 of 100


building tree 8 of 100
building tree 9 of 100building tree 10 of 100building tree 11 of 100


building tree 12 of 100
building tree 13 of 100building tree 14 of 100

building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   22.3s
```

```
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100
building tree 36 of 100
building tree 37 of 100
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100
building tree 50 of 100
building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
building tree 74 of 100
building tree 75 of 100
building tree 76 of 100
building tree 77 of 100
building tree 78 of 100
building tree 79 of 100
building tree 80 of 100
building tree 81 of 100
building tree 82 of 100
building tree 83 of 100
building tree 84 of 100
building tree 85 of 100
building tree 86 of 100
building tree 87 of 100
building tree 88 of 100
building tree 89 of 100
building tree 90 of 100
building tree 91 of 100
building tree 92 of 100
building tree 93 of 100
building tree 94 of 100
building tree 95 of 100
building tree 96 of 100
building tree 97 of 100
building tree 98 of 100
building tree 99 of 100
building tree 100 of 100

[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:  1.1min finished
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.4s finished
```

```
Out[300]: 0.002342528874500705
```

In [301]:

```
teste_X_9.log_vol.replace(teste_X_9.log_vol.values, pred_test_8, inplace=True)
teste_X_9['log_volTplus1']= Eval_month9.log_volTplus1

RelevantFeatures_month_10= RelevantFeatures_month_9.append(teste_X_9, ignore_index=True, sort=False)


train_X_10= RelevantFeatures_month_10.drop('log_volTplus1', axis=1)
train_y_10= RelevantFeatures_month_10.log_volTplus1
teste_X_10=Eval_month10.drop('log_volTplus1', axis=1)
teste_y_10=Eval_month10.log_volTplus1

modelRF.fit(train_X_10, train_y_10)
pred_test_10=modelRF.predict(teste_X_10)
error_10=mean_squared_error(teste_y_10, pred_test_10)
error_10
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 1 of 100
building tree 2 of 100
building tree 3 of 100
building tree 4 of 100
building tree 5 of 100
building tree 6 of 100building tree 7 of 100building tree 8 of 100


building tree 9 of 100
building tree 10 of 100
building tree 11 of 100
building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   19.2s
```

```
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100
building tree 36 of 100
building tree 37 of 100
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100
building tree 50 of 100
building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
building tree 74 of 100
building tree 75 of 100
building tree 76 of 100
building tree 77 of 100
building tree 78 of 100
building tree 79 of 100
building tree 80 of 100
building tree 81 of 100
building tree 82 of 100
building tree 83 of 100
building tree 84 of 100
building tree 85 of 100
building tree 86 of 100
building tree 87 of 100
building tree 88 of 100
building tree 89 of 100
building tree 90 of 100
building tree 91 of 100
building tree 92 of 100
building tree 93 of 100
building tree 94 of 100
building tree 95 of 100
building tree 96 of 100
building tree 97 of 100
building tree 98 of 100
building tree 99 of 100
building tree 100 of 100

[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:  1.1min finished
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.1s finished
```

```
Out[301]: 0.00251002427528707
```

In [302]:

```
teste_X_10.log_vol.replace(teste_X_10.log_vol.values, pred_test_10, inplace=True)
teste_X_10['log_volTplus1']= Eval_month10.log_volTplus1

RelevantFeatures_month_11= RelevantFeatures_month_10.append(teste_X_10, ignore_index=True, sort=False)


train_X_11= RelevantFeatures_month_11.drop('log_volTplus1', axis=1)
train_y_11= RelevantFeatures_month_11.log_volTplus1
teste_X_11=Eval_month11.drop('log_volTplus1', axis=1)
teste_y_11=Eval_month11.log_volTplus1

modelRF.fit(train_X_11, train_y_11)
pred_test_11=modelRF.predict(teste_X_11)
error_11=mean_squared_error(teste_y_11, pred_test_11)
error_11
```

```
[Parallel(n_jobs=-1)]: Using backend ThreadingBackend with 8 concurrent workers.

building tree 1 of 100building tree 2 of 100

building tree 3 of 100
building tree 4 of 100
building tree 5 of 100building tree 6 of 100

building tree 7 of 100
building tree 8 of 100
building tree 9 of 100
building tree 10 of 100
building tree 11 of 100
building tree 12 of 100
building tree 13 of 100
building tree 14 of 100
building tree 15 of 100
building tree 16 of 100
building tree 17 of 100
building tree 18 of 100
building tree 19 of 100
building tree 20 of 100
building tree 21 of 100
building tree 22 of 100
building tree 23 of 100
building tree 24 of 100
building tree 25 of 100
building tree 26 of 100
building tree 27 of 100
building tree 28 of 100
building tree 29 of 100
building tree 30 of 100
building tree 31 of 100
building tree 32 of 100

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   19.1s
```

```
building tree 33 of 100
building tree 34 of 100
building tree 35 of 100
building tree 36 of 100
building tree 37 of 100
building tree 38 of 100
building tree 39 of 100
building tree 40 of 100
building tree 41 of 100
building tree 42 of 100
building tree 43 of 100
building tree 44 of 100
building tree 45 of 100
building tree 46 of 100
building tree 47 of 100
building tree 48 of 100
building tree 49 of 100
building tree 50 of 100
building tree 51 of 100
building tree 52 of 100
building tree 53 of 100
building tree 54 of 100
building tree 55 of 100
building tree 56 of 100
building tree 57 of 100
building tree 58 of 100
building tree 59 of 100
building tree 60 of 100
building tree 61 of 100
building tree 62 of 100
building tree 63 of 100
building tree 64 of 100
building tree 65 of 100
building tree 66 of 100
building tree 67 of 100
building tree 68 of 100
building tree 69 of 100
building tree 70 of 100
building tree 71 of 100
building tree 72 of 100
building tree 73 of 100
building tree 74 of 100
building tree 75 of 100
building tree 76 of 100
building tree 77 of 100
building tree 78 of 100
building tree 79 of 100
building tree 80 of 100
building tree 81 of 100
building tree 82 of 100
building tree 83 of 100
building tree 84 of 100
building tree 85 of 100
building tree 86 of 100
building tree 87 of 100
building tree 88 of 100
building tree 89 of 100
building tree 90 of 100
building tree 91 of 100
building tree 92 of 100
building tree 93 of 100
building tree 94 of 100
building tree 95 of 100
building tree 96 of 100
building tree 97 of 100
building tree 98 of 100
building tree 99 of 100
building tree 100 of 100

[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed:  1.1min finished
[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.0s finished
```

```
Out[302]: 0.005438937080671865 5
```

Mean prediction error (mse) for the next 12 months

In [306]:

```
#erro_total=error_0+error_1+error_2+error_3+error_4+error_5+error_6+error_7+error_8+error_9+error_10+error_11

erro_total/12
```

Out[306]:

```
0.008249459935332158
```

In [ ]:

```
```