SUPSI

ANALISI TEMPO ATMOSFERICO DI SZEGED

Nicolas Gregori

Progetto di Machine Learning

7 Maggio 2021

SUPSI

Descrizione del Dataset

- Raccolta dati parametri atmosferici
- Avvenuta tra il 2006 e il 2016 in una città ungherese
- Contenuto:
  - 12 features
  - 96456 righe

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Loud Cover	Pressure (millibars)	Daily Summary
0	2006-04-01 00:00:00.000 +0200	Partly Cloudy	rain	9.472222	7.388889	0.89	14.1197	251.0	15.8263	0.0	1015.13	Partly cloudy throughout the day.
1	2006-04-01 01:00:00.000 +0200	Partly Cloudy	rain	9.355556	7.227778	0.86	14.2646	259.0	15.8263	0.0	1015.63	Partly cloudy throughout the day.
2	2006-04-01 02:00:00.000 +0200	Mostly Cloudy	rain	9.377778	9.377778	0.89	3.9284	204.0	14.9569	0.0	1015.94	Partly cloudy throughout the day.
3	2006-04-01 03:00:00.000 +0200	Partly Cloudy	rain	8.288889	5.944444	0.83	14.1036	269.0	15.8263	0.0	1016.41	Partly cloudy throughout the day.
4	2006-04-01 04:00:00.000 +0200	Mostly Cloudy	rain	8.755556	6.977778	0.83	11.0446	259.0	15.8263	0.0	1016.51	Partly cloudy throughout the day.

SUPSI

Preprocessing dei dati

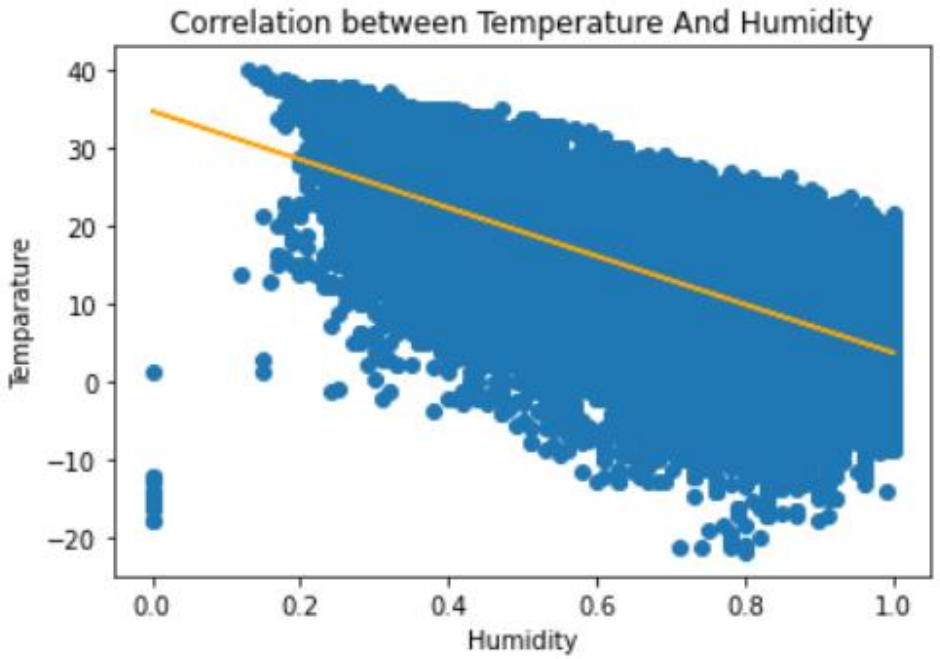
- Preprocessing dei dati quasi nullo
- Eliminazione delle colonne poco rilevanti
- Rimpiazzamento dei valori NaN delle features
- Splitting del dataset:
  - 60% train dataset
  - 40% test dataset

SUPSI

Temperatura - Umidità

- Regressione Lineare
  - Feature:
    - Temperatura
  - Target:
    - Umidità

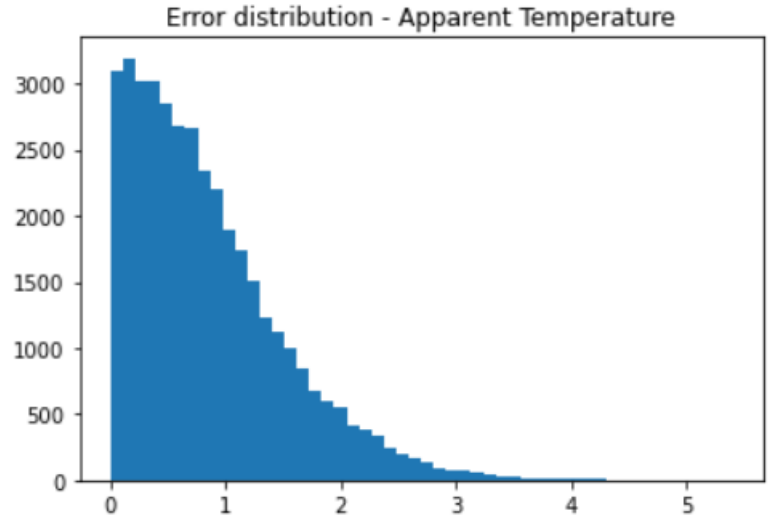
- R^2 test score: ~ 40%
- RMSE test score: ~ 7.39
- Regressore non molto buona



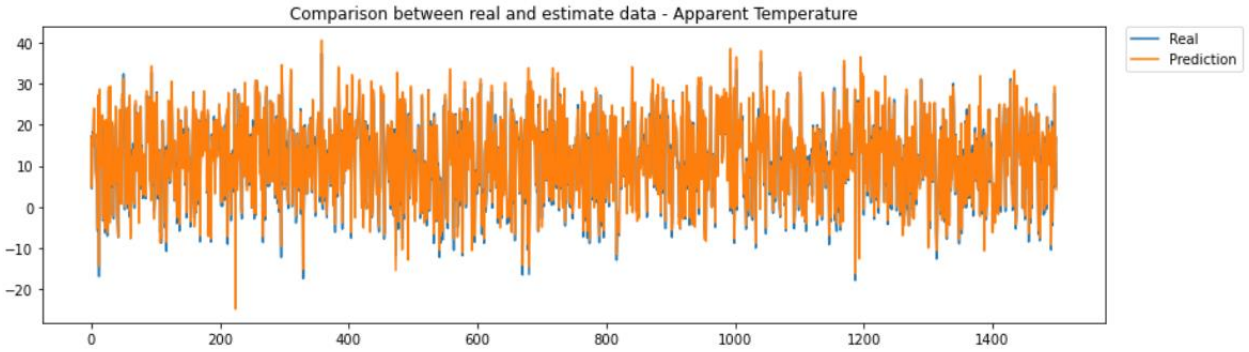
SUPSI

Temperatura Apparente

- Regressione Multipla
  - Features:
    - Temperatura
    - Umidità
    - Velocità del vento
  - Target:
    - Temperatura Apparente



- R^2 test score: ~98%
- RMSE test score: ~1.08
- Regressore molto buono



SUPSI

Pressione Atmosferica

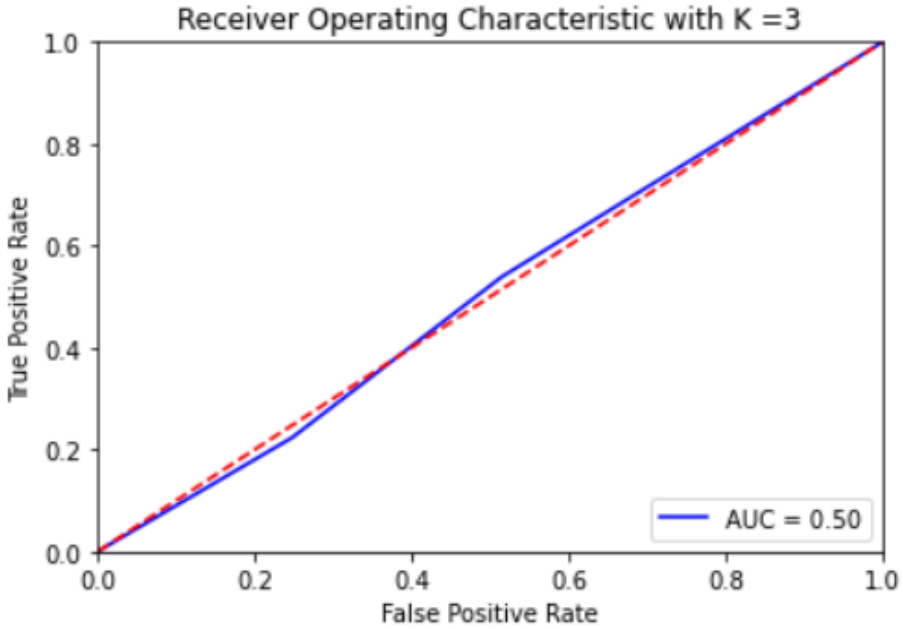
- Regressioni Ridge, Lasso e Elastic Net via GridSearchCV
  - Feature:
    - Tutti gli attributi target eccetto uno
  - Target:
    - Pressione atmosferica

- Alpha: [1e-15, 1e-10, 1e-08, 0.0001, 0.001, 0.01, 1, 5, 10]}
- Best RMSE (Elastic Net) : ~116.29

SUPSI

Tipo di precipitazione

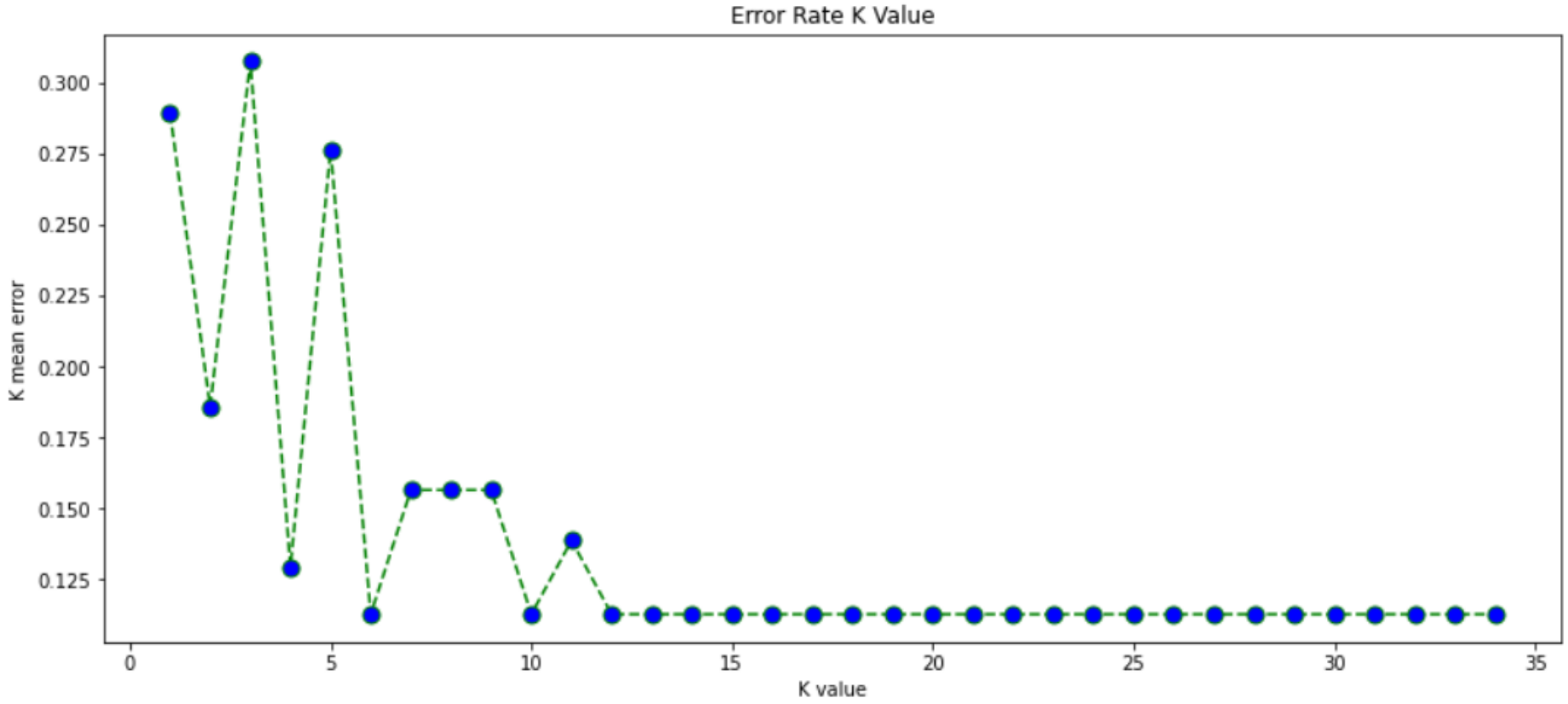
- Classificazione KNeighbors
  - Features:
    - Umidità
  - Target:
    - Tipo di Precipitazione




```
Confusion matrix:
[[25616  8435]
 [ 3358   966]]
Classification report:
              precision    recall  f1-score   support

        rain       0.88      0.75      0.81     34051
        snow       0.10      0.22      0.14      4324

    accuracy                           0.69     38375
   macro avg       0.49      0.49      0.48     38375
weighted avg       0.80      0.69      0.74     38375

Accuracy: 0.8252508143322476
```

SUPSI

Tipo di precipitazione (2)



SUPSI

Tipo di precipitazione (3)

- SVM tramite GridSearchCV
  - Linear
  - RBF

```
Params combination:
 [{'C': 1, 'kernel': 'linear'}, {'C': 5, 'kernel': 'linear'}, {'C': 10, 'kernel': 'linear'}]
Avg accuracy per combination:
 [0.88902208 0.88902208 0.88902208]
Best combination:
 {'C': 1, 'kernel': 'linear'}
Avg accuracy of best combination: 0.889022080922847
```

```
Params combination:
 [{'C': 1, 'gamma': 0.1, 'kernel': 'rbf'}, {'C': 1, 'gamma': 0.01, 'kernel': 'rbf'}, {'C': 5, 'gamma': 0.1, 'kernel':
'rbf'}, {'C': 5, 'gamma': 0.01, 'kernel': 'rbf'}, {'C': 10, 'gamma': 0.1, 'kernel': 'rbf'}, {'C': 10, 'gamma': 0.01,
'kernel': 'rbf'}]
Avg accuracy per combination:
 [0.88902208 0.88902208 0.88902208 0.88902208 0.88902208 0.88902208]
Best combination:
 {'C': 1, 'gamma': 0.1, 'kernel': 'rbf'}
Avg accuracy of best combination: 0.889022080922847
```