Previsão do número de visitantes em restaurantes Kaggle Competition

Cleilton C. D. da Silva cleiltoncarlos@copin.ufcg.edu.br

Por que a previsão?

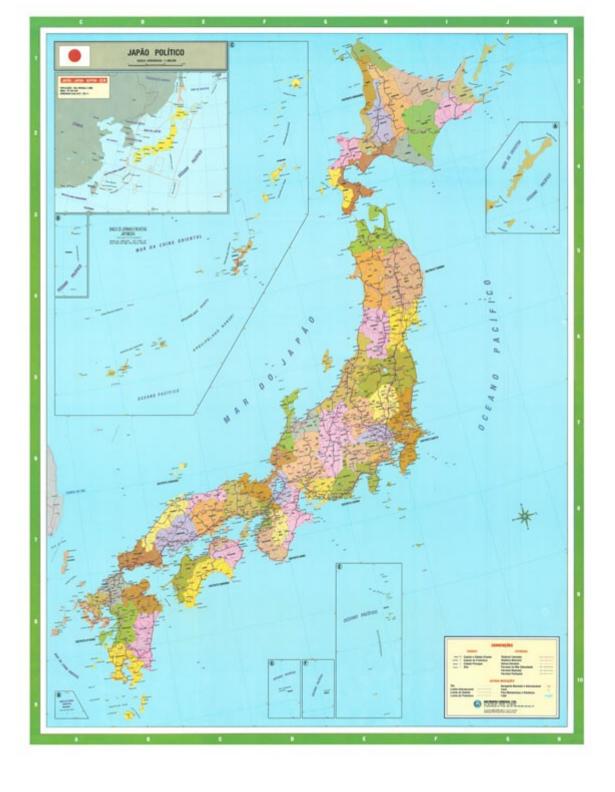


Por que a previsão?

- Compra de ingredientes
- Alocação de funcionários

Como prever número de clientes em restaurantes?

- O clima
- Feriado
- Final de semana
- Reservas em sites de compras
- Promoção



Experimento

- Validação Cruzada 2-fold
- Regressão Ridge e Lasso (sklearn)
- scoring="neg_mean_squared_log_error"
- Dados
 - Hot Pepper Gourmet (hpg)
 - AirREGI / Restaurant Board (air)

Experimento



Modelo 1 (features)

```
reserve_visitors
holiday_flg
weekend
```

(Utilizado somente dados de Air database)

Modelo 2 (features)

```
reserve_visitors
holiday_flg
weekend
from_air_database
from_hpg_database
```

(Utilizado dados de Air e HPG database)

Modelo 3 (features)

```
reserve_visitors
holiday_flg
weekend
```

(Inclusão de dados em que não houve reserva)

Modelo 4 (features)

Fatores:

- General information (3)
- Weather information (4)
- Restaurant information (6)

Interações

- Weather * Date (1)
- Reserve * Weather (2)
- Date * Restaurant (13)
- Reserve * Restaurant (4)

(Preenchimento de NaN baseado na média do dia da região)

Modelo 5 (features)

General information:

- reserve_visitors
- holiday_flg
- weekend

Weather information:

- t_media
- daylight_hours
- ave_cloud_cover

(Preenchimento de NaN baseado apenas na média do distrito)

Modelo 6 (features)

General information:

- reserve_visitors
- holiday_flg
- weekend

Weather information:

- t_media
- daylight_hours
- ave_cloud_cover

(Preenchimento de NaN baseado na média do distrito, considerando o ano e o mês.)

Modelo 7 (features)

Fatores

- General information (3)
- Restaurant information (11)

Interações

- Date * Restaurant (10)
- Reserve * Restaurant (5)

Model	Training RMLSE (Ridge)	Training RMLSE (Lasso)	Test RMLSE (Ridge)
1	0.591	0.591	-
2	0.656	0.656	-
3	0.829	0.829	-
4	0.810	0.812	0.830
5	0.831	0.832	0.852
6	0.829	0.830	0.847
7	0.813	0.815	0.824

Model	Training RMLSE (Ridge)	Training RMLSE (Lasso)	Test RMLSE (Ridge)
1	0.591	0.591	-
2	0.656	0.656	-
3	0.829	0.829	-
4	0.810	0.812	0.830
5	0.831	0.832	0.852
6	0.829	0.830	0.847
7	0.813	0.815	0.824

Model	Training RMLSE (Ridge)	Training RMLSE (Lasso)	Test RMLSE (Ridge)
1	0.591	0.591	-
2	0.656	0.656	-
3	0.829	0.829	-
4	0.810	0.812	0.830
5	0.831	0.832	0.852
6	0.829	0.830	0.847
7	0.813	0.815	0.824

Model	Training RMLSE (Ridge)	Training RMLSE (Lasso)	Test RMLSE (Ridge)
1	0.591	0.591	-
2	0.656	0.656	-
3	0.829	0.829	-
4	0.810	0.812	0.830
5	0.831	0.832	0.852
6	0.829	0.830	0.847
7	0.813	0.815	0.824

Conclusões

- Tratamento de NaNs
- Seleção de features
- Features com mais influência:
 - weekend
 - reserve_visitors
 - holiday_flg
 - is_dining_bar
 - visitors * italian_french