

Atualização do projeto PREPI:

Colab:

https://drive.google.com/file/d/1_3n2fHOkIb0A4VduvkUNXMYV0SFietxN/view?usp=sharing

Features:

<https://docs.google.com/spreadsheets/d/165aNT9qv-RGsCPG1SfyfZOSV0H3dkhxGjod95cqi5h8/edit?usp=sharing>

Random Forest

Evolução no demonstrativo de métricas. Random Forest continua ainda a ser um dos melhores modelos avaliados. Apesar disso, percebe-se que a precisão das previsões relacionadas a resultados positivos (became_client) ainda está bem abaixo.

```
Acc Treino: 91.11094826322731
Acc Teste: 88.90517367726807
=====
F1 Treino: 90.14782484138367
F1 Teste: 86.75669398725464
=====
Confusion Matrix
[[5675 198]
 [ 603 347]]
=====
Classification Report
      precision    recall  f1-score   support

     0       0.90      0.97      0.93     5873
     1       0.64      0.37      0.46      950

   accuracy      0.88      0.88      0.88     6823
  macro avg       0.77      0.67      0.70     6823
 weighted avg       0.87      0.88      0.87     6823

Precision: Percentage of correct positive predictions relative to total positive predictions
Recall: Percentage of correct positive predictions relative to total actual positives
F1 Score: A weighted harmonic mean of precision and recall. The closer to 1, the better the model
```

MLP Classifier

Tentativa de aplicação de modelos mais complexos, como o Multi-Layer Perception Classifier. Não apresenta resultados muito mais significativos que o RF.

```
Acc Treino: 90.1106551370365
Acc Teste: 88.90517367726807
=====
F1 Treino: 88.81557665557006
F1 Teste: 86.75669398725464
=====
Confusion Matrix
[[5715 158]
 [ 614 336]]
=====
Classification Report
      precision    recall  f1-score   support

     0       0.90      0.97      0.94     5873
     1       0.68      0.35      0.47      950

   accuracy      0.89      0.89      0.89     6823
  macro avg       0.79      0.66      0.70     6823
 weighted avg       0.87      0.89      0.87     6823

Precision: Percentage of correct positive predictions relative to total positive predictions
Recall: Percentage of correct positive predictions relative to total actual positives
F1 Score: A weighted harmonic mean of precision and recall. The closer to 1, the better the model
```

Feature importance

Para entender o que o modelo está levando em consideração para tomada de decisão, foi aplicado uma avaliação de Feature Importance. Este método, avalia para um índice específico, qual o peso das features para o resultado encontrado.

```
1 from lime import lime_tabular
2 import random

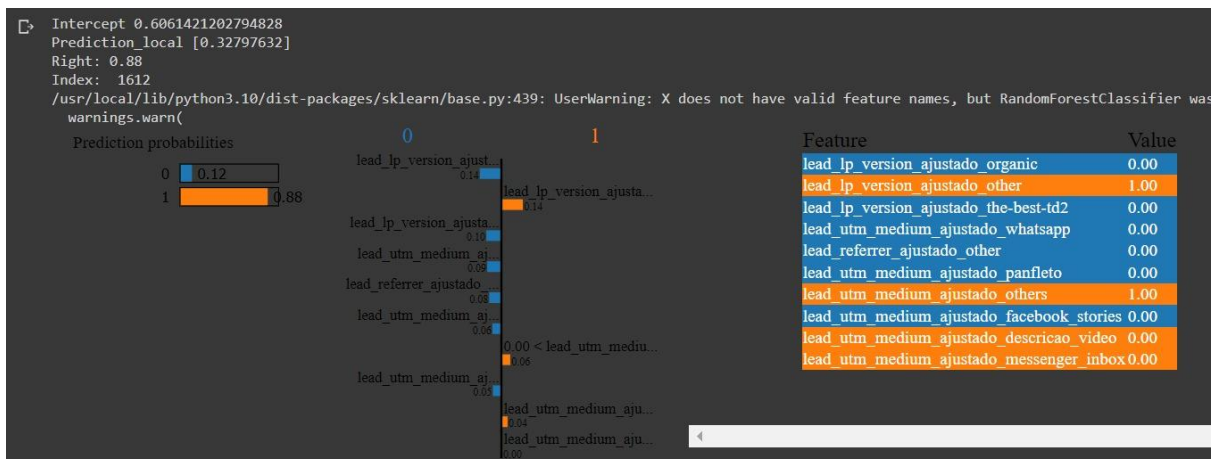
1 # Preparing the object to explain
2 explainer = lime_tabular.LimeTabularExplainer(np.array(X_train), mode = 'classification',
3                                               feature_names = X_train.columns,
4                                               verbose=True)
5 explainer

lime.lime_tabular.LimeTabularExplainer at 0x7fb807d4c7f0>

1 # Chosing a random index
2 idx = random.randint(1, len(X_test))
3
4 # Explaining the the specific object by index (iloc)
5 exp = explainer.explain_instance(X_test.iloc[2694], model_rf.predict_proba)
6
7 print('Index: ',idx)
8 exp.show_in_notebook(show_table=True)
```

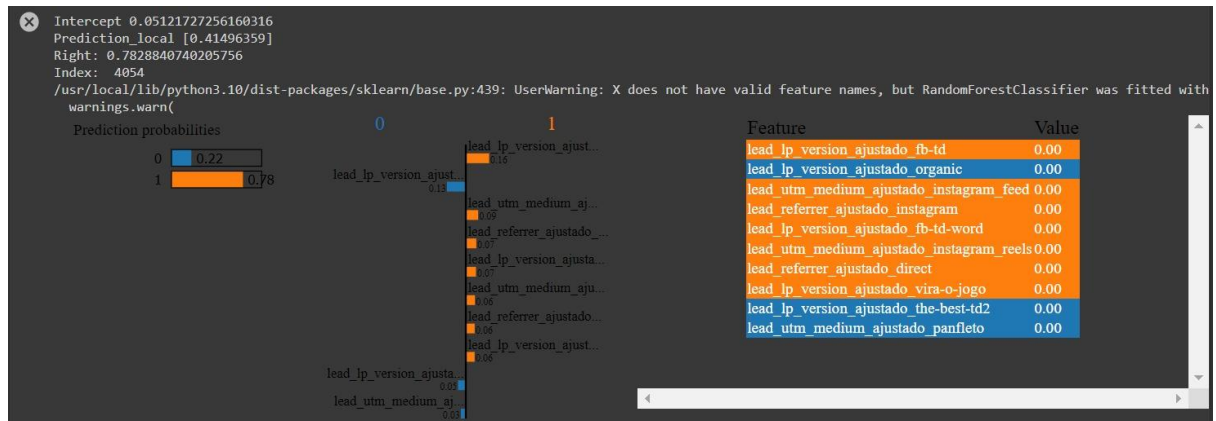
Utilizando um índice de um lead que o modelo classificou como “virou cliente”, percebeu-se que muitas features significativas eram de classificação “outros”, ou seja, não traziam uma informação relevante.

Com others



Foi realizado a remoção destas colunas do dataset e reaplicados os modelos, trazendo resultados mais específicos, explicando melhor a tomada de decisão.

Sem others



Próximas etapas:

1. Enriquecer o modelo com outras features e avaliar resultados do modelo