

# RM 357053 - WILSON ROBERTO DE MELO

June 30, 2025

## 1 Case QuantumFinance - Disciplina NLP - Classificador de chamados

**Atenção:** - Leia com atenção o descritivo do trabalho e as orientações do template. - O trabalho deve ser entregue respeitando a estrutura do arquivo de template em notebook “Template\_Trabalho\_Final\_NLP.ipynb” e compactado no formato .zip. Apenas um arquivo no formato .ipynb deve ser entregue consolidando todo o trabalho.

**Participantes (RM - NOME):** # RM 357053 - WILSON ROBERTO DE MELO

### 1.1 ### Crie um classificador de chamados aplicando técnicas de PLN

A **QuantumFinance** tem um canal de atendimento via chat e precisa classificar os assuntos dos atendimentos para melhorar as tratativas dos chamados dos clientes. O canal recebe textos abertos dos clientes relatando o problema e/ou dúvida e depois é direcionado para alguma área especialista no assunto para uma melhor tratativa.

1. Crie ao menos um modelo classificador de assuntos aplicando técnicas de NLP (PLN), Vetorização (n-grama + métrica) e modelo supervisionado, que consiga classificar através de um texto o assunto conforme disponível na base de dados [1] para treinamento e validação do seu modelo.

O modelo precisa atingir um score na **métrica F1 Score superior a 75%**. Utilize o dataset [1] para treinar e testar o modelo, separe o dataset em duas amostras (75% para treinamento e 25% para teste com o `random_state` igual a 42).

2. Utilizar ao menos uma aplicação de modelos com Embeddings usando Word2Vec e/ou LLM's para criar o modelo classificador com os critérios do item 1. Não é necessário implementar aplicações usando serviços de API da OpenAI ou outros por exemplo.

Fique à vontade para testar e explorar as técnicas de pré-processamento, abordagens de NLP, algoritmos e bibliotecas, mas explique e justifique as suas decisões durante o desenvolvimento.

### Composição da nota:

**50%** - Demonstrações das aplicações das técnicas de PLN (regras, pré-processamentos, tratamentos, variedade de modelos aplicados, aplicações de GenIA, organização do pipeline, etc.)

**50%** - Baseado na performance (score) obtida com a amostra de teste no pipeline do modelo campeão (validar com a Métrica F1 Score). **Separar o pipeline completo do modelo campeão conforme template.**

O trabalho poderá ser feito em grupo de 2 até 4 pessoas (mesmo grupo do Startup One) e trabalhos iguais serão descontado nota e passível de reprovação.

[1] = [https://dados-ml-pln.s3.sa-east-1.amazonaws.com/tickets\\_reclamacoes\\_classificados.csv](https://dados-ml-pln.s3.sa-east-1.amazonaws.com/tickets_reclamacoes_classificados.csv)

**F1 Score** com average='weighted'

```
[1]: # CARREGANDO O DATA FRAME
import pandas as pd
df = pd.read_csv('https://dados-ml-pln.s3.sa-east-1.amazonaws.com/
↳tickets_reclamacoes_classificados.csv', delimiter=';')

# Façam o download do arquivo e utilizem localmente durante os testes
```

```
-----
ModuleNotFoundError                                Traceback (most recent call last)
Cell In[1], line 2
      1 # CARREGANDO O DATA FRAME
----> 2 import pandas as pd
      3 df = pd.read_csv('https://dados-ml-pln.s3.sa-east-1.amazonaws.com/
↳tickets_reclamacoes_classificados.csv', delimiter=';')
      5 # Façam o download do arquivo e utilizem localmente durante os testes

ModuleNotFoundError: No module named 'pandas'
```

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21072 entries, 0 to 21071
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id_reclamacao          21072 non-null  int64
1   data_abertura          21072 non-null  object
2   categoria              21072 non-null  object
3   descricao_reclamacao  21072 non-null  object
dtypes: int64(1), object(3)
memory usage: 658.6+ KB
```

Bom desenvolvimento!

### ###Area de desenvolvimento e validações

Faça aqui as demonstrações das aplicações das técnicas de PLN (regras, pré-processamentos, tratamentos, variedade de modelos aplicados, organização do pipeline, etc.)

Fique à vontade para testar e explorar as técnicas de pré-processamento, abordagens de NLP, algoritmos e bibliotecas, mas explique e justifique as suas decisões durante o desenvolvimento.

```
[ ]:
```

[ ]:

[ ]:

[ ]:

### ###Validação do professor

Consolidar apenas os scripts do seu **modelo campeão**, desde o carregamento do dataframe, separação das amostras, tratamentos utilizados (funções, limpezas, etc.), criação dos objetos de vetorização dos textos e modelo treinado e outras implementações utilizadas no processo de desenvolvimento do modelo.

O modelo precisar atingir um score na métrica F1 Score superior a 75%.

**Atenção:** - Implemente aqui apenas os scripts que fazem parte do modelo campeão.  
- Execute o pipeline do modelo campeão completamente para garantir que não terá erros no script.

[ ]:

[ ]:

[ ]:

[ ]: