

# DETECTION OF BANKING ORGANIZATIONS USING NAMED ENTITY RECOGNITION

*Rubens José Teixeira Machado Neto*

Universidade Federal da Bahia

## **Abbreviated abstract:**

Named Entity Recognition(NER) is a technique used to classify entities from text in Natural Language Processing(NLP) which is capable of extracting and classifying predefined entities. The overall goal of this project is to produce Natural Language Processing models for identifying banking entities in Portuguese language texts using a pretrained BERT model.

## **Related publications:**

– Heck, Amabyle Rabeche. *Processamento de linguagem natural aplicado a reconhecimento de entidades nomeadas em textos legais em português brasileiro.*

MS thesis. Florianópolis, SC., 2022.



[Rubens.Teixeira@outlook.com.br](mailto:Rubens.Teixeira@outlook.com.br)



4th Conference on  
**Statistics and  
Data Science**  
Salvador, Brazil (online)  
December 1-3, 2022

# Problem and Data

## Problem:

- Identify, categorize and extract banking entities in unstructured text.

## Datasets:

- NER Noticias – Dataset ready from Github with general News already labelled with “Pessoa, Org, Pub and LOC” (187 News)
- Noticias Bancarias – Dataset scrapped and built from Google Search and Wikipedia and labelled with “Banco” using ten banks from Brazil (50 text – 40 News + 10 bank summaries from Wikipedia).

The overall goal of this project is to produce Natural Language Processing models for identifying banking entities in Portuguese language texts using a pretrained BERT model.

```
{ "id": 348, "text": "Operação da PF investiga suspeita de fraude na compra de testes para Covid-19 em R$ 10 milhões. Há fortes indícios, segundo a PF, de irregularidades no chamamento público para compra dos testes e superfaturamento no valor de cada unidade adquirida. O valor total da contratação investigada passa dos R$ 10 milhões.", "meta": {}, "annotation_approver": null, "labels": [[12, 14, "PUB"], [81, 83, "PUB"], [115, 117, "PUB"]] }
```

<p>Inter, antigo Banco Inter e Banco Intermedium, é uma empresa de tecnologia brasileira, fundada em 1994 e sediada em Belo Horizonte, Minas Gerais. Caracteriza-se por ser um dos primeiros bancos digitais do Brasil, e pela ausência de taxas sobre serviços básicos. Atualmente, além de serviços financeiros, oferece também serviços não financeiros, se posicionando como um Super App.</p>	<p>O Banco Inter anunciou, na última semana, o lançamento do Invest Now, Pay Later, produto financeiro que permite investir usando o cartão de crédito. O novo recurso oferece a possibilidade de investir em uma Letra de Crédito Imobiliário (LCI), com liquidez de 180 dias, na faixa de R\$ 500 a R\$ 5 mil. Na prática, o cliente realiza o investimento como uma compra em seu cartão de crédito e paga por esse produto depois da aquisição, no momento do vencimento da fatura do cartão. Os aportes são isentos de descontos no imposto de renda e de quaisquer outras taxas e recebem a proteção do Fundo Garantidor de Crédito (FGC). Para que o cliente possa resgatar o investimento na data estipulada, é necessário pagar de forma integral a fatura do cartão de crédito que inclui o aporte do produto. Em caso de inadimplência do cartão ou de pagamento parcial da fatura, o cliente não poderá ter acesso à rentabilidade do investimento. O lançamento ainda prevê outras possibilidades de investimentos a curto ou médio prazo, que permitam aportes em mais produtos, pagamento parcelado e diferentes prazos para retirada do dinheiro.</p>	<p>O Inter divulgou na última segunda-feira (15) lucro líquido de R\$ 16 milhões para o segundo trimestre, ante prejuízo de R\$ 30 milhões obtidos no mesmo período do ano passado. O resultado veio após a companhia ter elevado a receita líquida de serviços em 91% no segundo trimestre em relação a um marketplace, apurou um lucro líquido de quase 28% nas vendas brutas pelo conceito GMV, que atingiu 990 milhões de reais no segundo trimestre. O número de clientes ativos subiu 51,7%, para 10,7 milhões, enquanto o custo de servir cada um deles recuou 4,5%, para R\$ 14. Mas o custo de aquisição de cliente subiu quase 18%, para R\$ 32. A rentabilidade sobre o patrimônio encerrou o trimestre positiva em 0,8%, ante 2% negativos um ano antes. Fundador e presidente do conselho da MRV Engenharia, o empresário Rubens Menin é também controlador do Inter, da Log Commercial Properties e da CNN Brasil.</p>	<p>O Banco Inter inovou e lançou uma novidade para os clientes que gostam de investir seu dinheiro. A instituição criou a "Invest Now, Pay Later" (em português: invista agora, pague depois), ferramenta que promete modificar a maneira como seus usuários investem. O investidor poderá fazer uma aplicação financeira para pagar no cartão de crédito, como se estivesse fazendo uma compra no cartão. Na data de vencimento da fatura, é só pagar a dívida normalmente. O primeiro produto disponível é uma Letra de Crédito Imobiliário (LCI), com liquidez de 180 dias, na qual é possível aplicar entre R\$ 500 e R\$ 5 mil. Nos próximos meses, o banco digital planeja acrescentar novas opções de investimento para ampliar o Invest Now, Pay Later. Os aportes por meio do recurso são isentos de imposto de Renda e de qualquer outra taxa, além de contarem com proteção do Fundo Garantidor de Crédito (FGC). "Essa novidade chega para mexer com o mercado, incentivando os clientes do Inter a investir com a vantagem de ter um período maior de rentabilidade em relação à data de pagamento do investimento", disse Felipe Bottino, diretor da Inter Invest. O novo produto tem apenas uma condição especial. Para resgatar a aplicação na data fixada, o cliente deverá ter realizado o pagamento integral da fatura do cartão de crédito no qual o aporte do ativo está incluído. No caso de inadimplência, ele ficará impedido de receber os rendimentos.</p>	<p>O Banco Inter anunciou que está lançando mais um produto financeiro para seus clientes: o Invest Now, Pay Later. Na prática, investidores vão poder realizar investimentos e só pagar os aportes na data de vencimento da fatura do cartão de crédito. A proposta é que, no futuro, seja permitido viabilizar outras parcelamentos dos valores investidos. Neste primeiro momento, o Invest Now, Pay Later vai oferecer aos clientes do Inter a possibilidade de investir em uma Letra de Crédito Imobiliário (LCI), com liquidez de 180 dias, na faixa de valor entre R\$ 500 e R\$ 5 mil. Como próximos passos, o banco pretende viabilizar outras possibilidades de investimentos. Os aportes em investimentos por meio do Invest Now, Pay Later são isentos de imposto de renda e de quaisquer outras taxas, e recebem a proteção do Fundo Garantidor de Crédito (FGC), diz o Inter. "Essa novidade chega para mexer com o mercado, incentivando os clientes do Inter a investir com a vantagem de ter um período maior de rentabilidade em relação à data de pagamento do investimento", afirma Felipe Bottino, diretor da Inter Invest. A única exigência do banco para a utilização do novo produto financeiro será no hora de resgatar o investimento. Para fazer isso na data estipulada, é preciso ter realizado o pagamento integral da fatura do cartão de crédito no qual o aporte não é ativo. Em caso de inadimplência ou pagamento parcial da fatura, o cliente não poderá ter acesso à rentabilidade do investimento.</p>
--	---	--	--	---



# Methods

Pre-processing dataset Noticias Bancarias:

- Choose 10 Banks from Brazil;
- Scrape data from Google Search and get summary of each Bank from Wikipedia API;
- Label all banks in this dataset with “Banco” using Docanno Platform.

Pre-processing dataset NER Noticia:

- Already Labelled data;

Pre-trained Model:

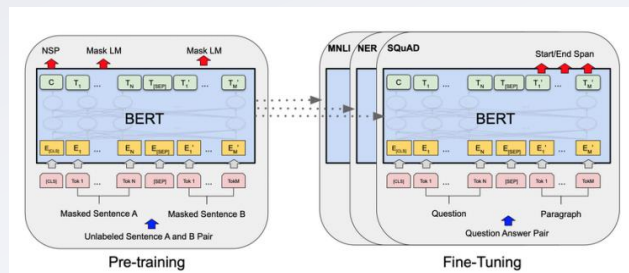
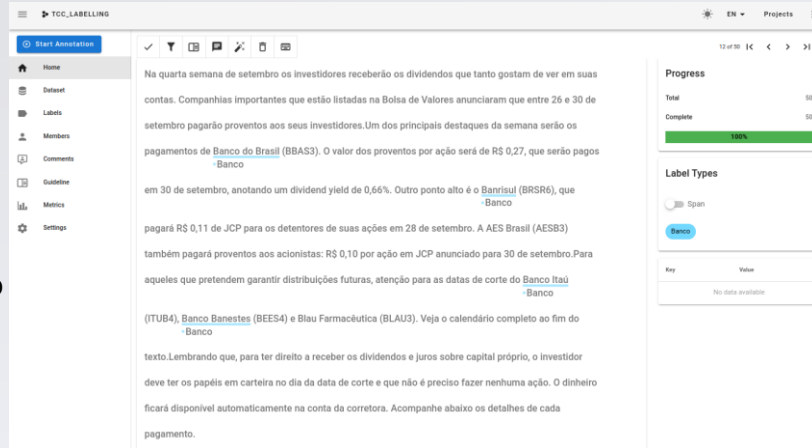
- Get a pretrained model called Bertimbau Large;

Fine Tuning Training:

- Using Noticias Bancarias and Ner Noticias to fine-tuning the model for NER task;

Model Assessment:

Take the model trained and test using Tweets got from Twitter;



# Results and Conclusions

Tabela 1 – Métricas

	Acurácia	Precisão	Recall	F1-Score
Modelo - Notícias Bancárias	0,984	0,870	0,940	0.901
Modelo - NER Notícias	0,984	0,867	0,921	0.893

Tabela 2 – Métricas Twitter

	Acurácia	Precisão	Recall	F1-Score
Modelo - Notícias Bancárias	0,680	0,903	0,491	0,636
Modelo - NER Notícias	0,690	0,933	0,491	0,644

O Banco Inter é uma instituição financeira que atua no mercado de crédito e de investimentos.

```
{'entity_group': 'Banco', 'score': 0.7138555, 'word': 'Banco Inter', 'start': 2, 'end': 13}
```

Preciso sacar dinheiro no BB

```
{'entity_group': 'ORG', 'score': 0.90257627, 'word': 'BB', 'start': 26, 'end': 28}
```

- Good performance for both models, using their own dataset (splitted 20% validation and 80% for Training).
- Average performance using Tweets due to the informality of the vocabulary and sentence construction.
- The model “Noticias Bancarias” identify easily the Banks. However NER Notícias model needs another second step(filter) in order to catch only the Banking Organizations because their labelling catches all organizations.

Further studies:

- Get and train a larger banking dataset;
- Use other pretrained model and tokenizers besides Bertimbau Large;
- Modify the hyperparameters from the model in order to achieve better results.

