

**Universidade de Brasília**  
**Departamento de Engenharia Elétrica**

**Trabalho Final - Previsão de resultados da eleição de 2022**  
**Noções de Inteligência Artificial**

Bruno Barros Xavier - 200015621  
Gabriel Francisco de Oliveira Castro - 202066571

## 1 Introdução

Ao explorar dados estatísticos relacionados a eventos, como eleições, a regressão linear oferece uma ferramenta valiosa para quantificar e prever padrões observados. Nas esferas políticas, onde a análise de dados desempenha um papel cada vez mais importante, a regressão linear emerge como uma abordagem poderosa para examinar a influência de diferentes fatores sobre os resultados eleitorais. Nesta análise, exploraremos como a regressão linear pode ser aplicada ao contexto das eleições, buscando entender como variáveis como gastos de renda, idade e outros fatores se relacionam e impactam nos resultados eleitorais.

Neste trabalho, criaremos uma rede neural capaz de calcular as probabilidades do eleitor pertencente ao universo da Universidade de Brasília (UnB) ter votado no LUIZ INÁCIO LULA DA SILVA (PT), JAIR MESSIAS BOLSONARO (PL) ou absteve-se. A abordagem para essa tarefa envolverá a implementação de uma rede neural com classificação multiclasse utilizando a função softmax e representação one-hot para os rótulos. Nesse sentido, a função softmax será responsável por atribuir probabilidades às diferentes classes possíveis.

## 2 Obtenção dos Dados

Para esse trabalho, usamos um banco de dados cedido pelo departamento de estatística da Universidade de Brasília. Os dados abrangem alguns parâmetros de informações, como a temporalidade dos eventos até as características socioeconômicas dos eleitores.

	Data	Pontuacao	Tipo_Coleta	Sexo	Fx_etaria	Fx_renda	Presidencia	Governo	Turno	Hora
0	2022-08-04 18:58:49 UTC	0	AAS sistemática	Feminino	19 a 24	1 a 2 SM	LUIZ INÁCIO LULA DA SILVA (PT)	Não sei/não quero responder	Noturno	18:58:49
1	2022-08-04 19:00:30 UTC	0	AAS sistemática	Masculino	25 a 29	5 a 10 SM	LUIZ INÁCIO LULA DA SILVA (PT)	Não sei/não quero responder	Noturno	19:00:29
2	2022-08-04 19:01:50 UTC	0	AAS sistemática	Masculino	19 a 24	5 a 10 SM	Não sei/ não quero responder	Não sei/não quero responder	Noturno	19:01:49
3	2022-08-04 19:02:40 UTC	0	AAS sistemática	Masculino	19 a 24	5 a 10 SM	LUIZ INÁCIO LULA DA SILVA (PT)	Leandro Grass (PV)	Noturno	19:02:39
4	2022-08-04 19:04:02 UTC	0	AAS sistemática	Masculino	19 a 24	mais de 10 SM	SIMONE TEBET (MDB)	Leandro Grass (PV)	Noturno	19:04:02
...	...	...	...	...	...	...	...	...	...	...
195	2022-08-01 16:06:58 UTC	0	AAS sistemática	Feminino	19 a 24	Até 1 SM	VERA LÚCIA (PSTU)	Não sei/ Não quero responder	Diurno	16:06:58
196	2022-08-02 08:46:31 UTC	0	AAS sistemática	Feminino	19 a 24	5 a 10 SM	CIRO GOMES (PDT)	Keka Bagno (PSOL)	Diurno	08:46:31
197	2022-07-26 11:20:43 UTC	0	AAS sistemática	Feminino	19 a 24	Sem Resposta	Não sei/ Não quero responder	Não sei/ Não quero responder	Diurno	11:20:43
198	2022-08-02 09:10:19 UTC	0	AAS sistemática	Feminino	19 a 24	mais de 10 SM	LUIZ INÁCIO LULA DA SILVA (PT)	Keka Bagno (PSOL)	Diurno	09:10:19
199	2022-08-02 08:55:09 UTC	0	AAS sistemática	Feminino	19 a 24	mais de 10 SM	LUIZ INÁCIO LULA DA SILVA (PT)	Branco	Diurno	08:55:09

Figura 1: Banco de dados com algumas colunas.

### 3 Tratamento dos Dados

Dentro do escopo deste projeto, diversos conjuntos de dados apresentam múltiplas entradas. Po-  
dendo na hora dificultar o treinamento e a análise da rede, já que a rede precisa lidar com multiplas  
opções, ocasionando um baixa taxa de acerto.

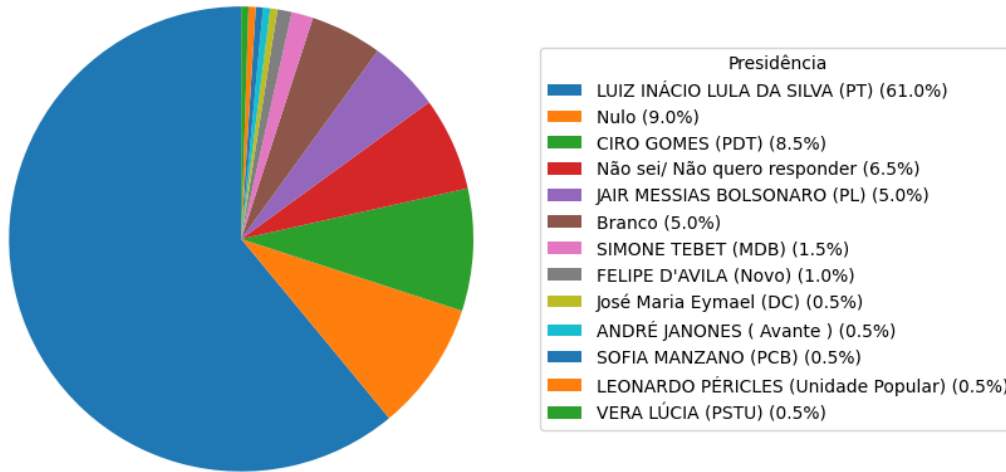


Figura 2: Distribuição de intenção de votos para presidente na UnB.

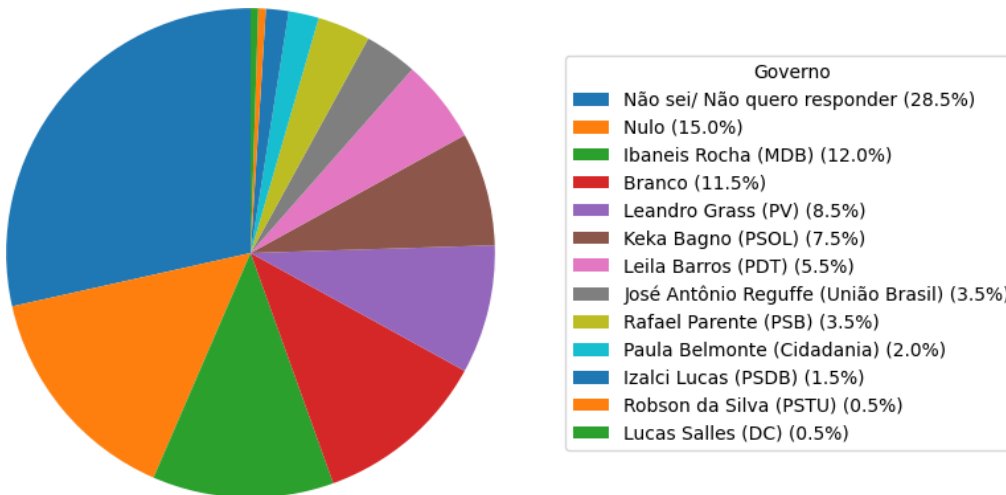


Figura 3: Distribuição de intenção de votos para governador na UnB.

Por isso, faz-se necessário a simplificação das seguintes informações:

- Presidente para apenas 3 classes: Luiz Inácio Lula da Silva, Jair Messias Bolsonaro ou Abs-  
tenção.
- Governador para apenas 5 classes: Direita, Esquerda, Centro-Direita, Centro-Esquerda ou  
Abstenção.
- Período para apenas 3 classes: Manhã, Tarde ou Noite.
- Turno para apenas 2 classes: Diurno ou Noite.

Outro problema foi a coleta manual de dados, muitas informações foram registradas de maneira incorreta. Como resultado, uma considerável partes dos dados são classificados de forma separada, por conta disso tiveram que ser corrigidas manualmente. Adicionalmente, algumas colunas, como "Data", "Pontuação" e "Tipo da Coleta", não agregam substancialmente à análise, e portanto foram removidas.

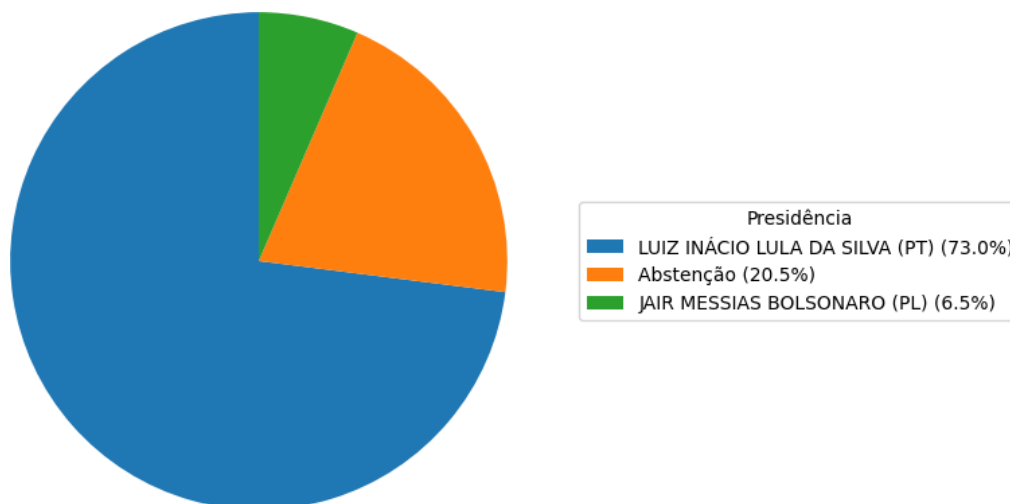


Figura 4: Distribuição simplificada das intenções de voto para presidente na UnB.

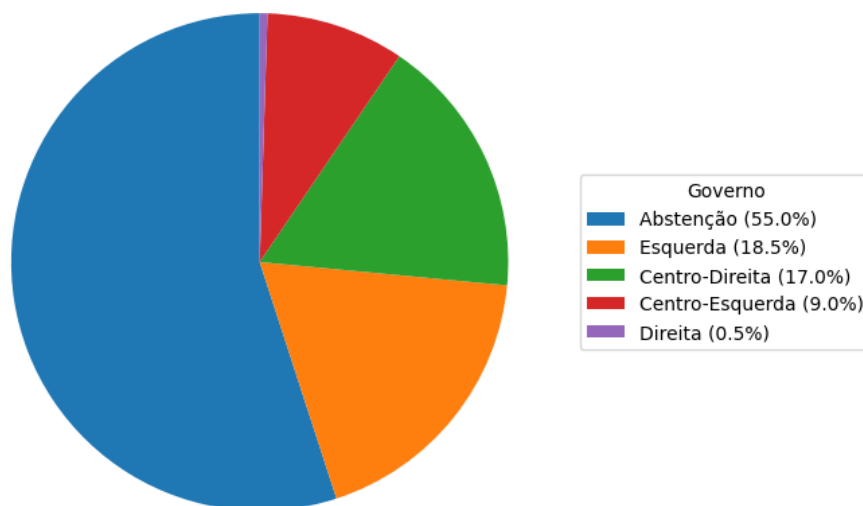


Figura 5: Distribuição simplificada das intenções de voto para governador na UnB.

Em relação a distribuição de votos para presidentes, podemos notar que a maioria dos eleitores do Universo da UnB (Figura 4) tinha a intenção de votar no candidato Lula, o que não é uma surpresa considerando o ambiente da faculdade, bem como o gráfico de distribuição de votos antes dos dados serem tratados. No entanto, o número significativo de eleitores que se abstiveram do voto levanta uma questão intrigante: será que esses eleitores realmente abstiveram-se ou optaram por não responder devido a algum receio?

	Sexo	Fx_etaria	Fx_renda	Presidencia	Governo	Turno	Periodo
0	0.0	1.0	1.0	1.0	0.0	1.0	0.0
1	1.0	2.0	2.0	1.0	0.0	1.0	0.0
2	1.0	1.0	2.0	0.0	0.0	1.0	0.0
3	1.0	1.0	2.0	1.0	1.0	1.0	0.0
4	1.0	1.0	3.0	1.0	1.0	1.0	0.0
...	...	...	...	...	...	...	...
195	0.0	1.0	5.0	1.0	0.0	2.0	2.0
196	0.0	1.0	2.0	1.0	1.0	2.0	1.0
197	0.0	1.0	0.0	0.0	0.0	2.0	1.0
198	0.0	1.0	3.0	1.0	1.0	2.0	1.0
199	0.0	1.0	3.0	1.0	0.0	2.0	1.0

Figura 6: Banco de dados com algumas colunas removidas e dados codificados de forma numérica.

## 4 Treinamento da Rede

Para otimizar os parâmetros da rede e ajustá-la de forma a minimizar a função de custo, foi utilizado algoritmo de otimização de Gradiente Descendente. Este algoritmo atua iterativamente, refinando os valores dos parâmetros na direção que reduz a discrepância entre as previsões do modelo e os valores reais. O objetivo é encontrar valores de parâmetros que resultem nas previsões mais precisas possível. Além da escolha criteriosa da função de otimização, outro aspecto crucial no treinamento de uma rede neural é a seleção adequada das funções de ativação. Com relação a isso, foi utilizado a função de ativação ReLu.

```
Epoch 1/20000
3/3 [=====] - 0s 90ms/step - loss: 1.1049 - accuracy: 0.3412
Epoch 2/20000
3/3 [=====] - 0s 14ms/step - loss: 1.0368 - accuracy: 0.7412
Epoch 3/20000
3/3 [=====] - 0s 13ms/step - loss: 0.9768 - accuracy: 0.7412
Epoch 4/20000
3/3 [=====] - 0s 13ms/step - loss: 0.9315 - accuracy: 0.7412
Epoch 5/20000
3/3 [=====] - 0s 14ms/step - loss: 0.8922 - accuracy: 0.7412
Epoch 6/20000
3/3 [=====] - 0s 13ms/step - loss: 0.8585 - accuracy: 0.7412
Epoch 7/20000
3/3 [=====] - 0s 13ms/step - loss: 0.8323 - accuracy: 0.7412
Epoch 8/20000
3/3 [=====] - 0s 13ms/step - loss: 0.8137 - accuracy: 0.7412
Epoch 9/20000
3/3 [=====] - 0s 13ms/step - loss: 0.8000 - accuracy: 0.7412
Epoch 10/20000
3/3 [=====] - 0s 13ms/step - loss: 0.7909 - accuracy: 0.7412
Epoch 11/20000
3/3 [=====] - 0s 13ms/step - loss: 0.7839 - accuracy: 0.7412
Epoch 12/20000
3/3 [=====] - 0s 15ms/step - loss: 0.7787 - accuracy: 0.7412
Epoch 13/20000
...
Epoch 19999/20000
3/3 [=====] - 0s 14ms/step - loss: 0.1842 - accuracy: 0.9118
Epoch 20000/20000
3/3 [=====] - 0s 14ms/step - loss: 0.1847 - accuracy: 0.9118
```

Nesse ponto, é interessante observar que devido ao número reduzido de dados, aproximadamente 200, optamos por configurar um número considerável de épocas com o objetivo de aprimorar a precisão do modelo desenvolvido e garantir que ele realmente aprendesse.

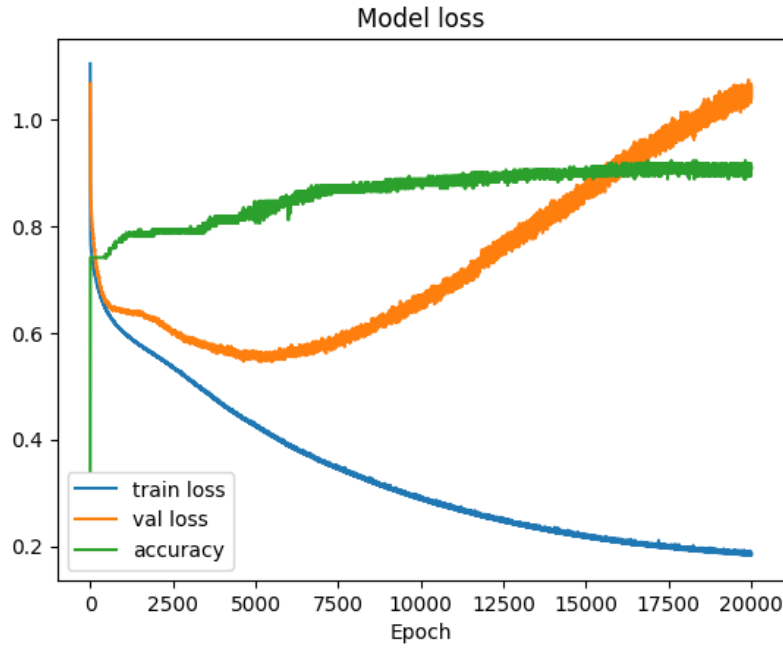


Figura 7: Perda e acurácia do modelo a cada época.

Plotando os resultados do treinamento, é possível ver uma minimização do erro e uma acurácia extremamente satisfatória de treinamento. No entanto, a partir de um momento existe uma grande maximização no erro de validação, podendo indicar um possível overfitting. É importante mencionar, quando o número de épocas é menor, como por exemplo, com 200 épocas, ocorre um fenômeno interessante. A acurácia do modelo permanece alta, por volta de 80%, mas há uma incerteza sobre se o modelo realmente aprendeu ou apenas identificou que ao admitir que a maioria das entradas fosse 'Lula', teria uma acurácia alta, uma vez que a maioria dos eleitores declarou intenção de voto no candidato do PT. Com isso em mente, optamos por aumentar drasticamente o número de épocas para garantir que o modelo tenha, de fato, aprendido, e obtivemos resultados bem satisfatórios.

Um ponto que chama a atenção. é a regressão linear poder funcionar bem quando há poucos dados disponíveis, enquanto MLPs geralmente requerem grandes volumes de dados para aprender representações eficazes. Tornando a regressão linear uma escolha mais sensata, quando o conjunto de dados é pequeno.

## 5 Resultados

Percebe-se como a probabilidade do eleitor votar em determinado candidato realmente reflete o esperado. Abaixo, apresentaremos de maneira mais clara três exemplos extremos no que se refere a classificação de 3 classes (Lula, Bolsonaro e Abstenção) nos quais a probabilidade apresentada está bem interessante.

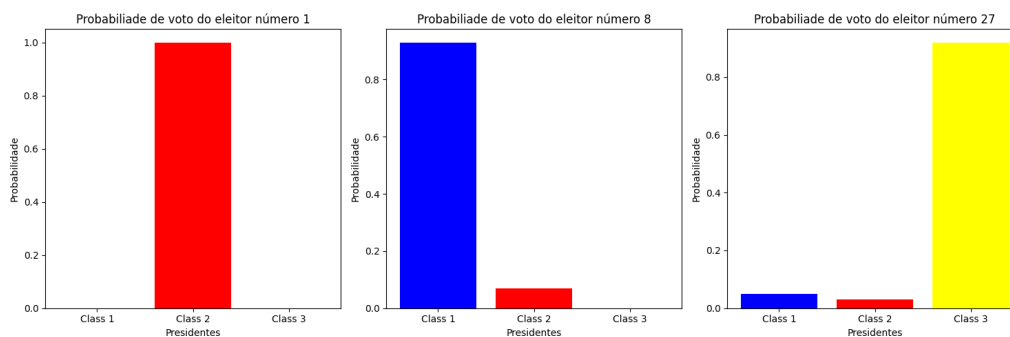


Figura 8: Três exemplos com a probabilidade de voto.

Estes três eleitores exibem maiores probabilidades de votar em Lula, Abstenção e Bolsonaro, respectivamente. Ao analisarmos quem eles declararam efetivamente como intenção de voto, notamos que as probabilidades se alinham perfeitamente com a escolha do respectivo eleitor. Isso confirma de maneira robusta que o modelo, de fato, aprendeu a partir dos dados disponíveis.

## 6 Conclusão

Ao longo do processo, destacamos a importância da análise exploratória de dados para compreender a natureza das informações disponíveis. A divisão apropriada dos dados em conjuntos de treinamento e teste permitiu a validação eficaz do desempenho do modelo em dados não vistos. Em síntese, os resultados obtidos revelam-se altamente satisfatórios, evidenciando uma correspondência significativa entre as probabilidades previstas pelo modelo e as escolhas efetivas dos eleitores. Esses resultados reforçam a eficácia da aplicação da rede neural na compreensão e previsão das preferências eleitorais, especialmente no contexto peculiar da Universidade de Brasília (UnB). A precisão do modelo ao capturar as nuances das preferências eleitorais na comunidade acadêmica sugere sua utilidade como ferramenta analítica.

## Referências

- [1] Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola (2023) Dive into Deep Learning, Cambridge University Press; 1st ed