



UNIVERSIDADE DE BRASÍLIA

FACULDADE DO GAMA

APRENDIZADO DE MÁQUINA - Turma 01 - 2025/1

Ratos da Câmara - grupo 09

Mini Trabalho 1:
Definição do problema e contextualização

Andre Ricardo Meyer de Melo - 231011097

Davi Rodrigues da Rocha - 211061618

Luiz Felipe Bessa Santos - 231011687

Tiago Antunes Balieiro - 231011838

Wesley Pedrosa dos Santos - 180029240

Brasília, DF

2025

1. Problema escolhido

a. Detectar gastos anômalos na cota parlamentar.

É possível identificar gastos atípicos ou suspeitos feitos por deputados com verba de gabinete?

b. Prever se um projeto de lei será aprovado.

Dado um novo projeto de lei, é possível prever se ele será aprovado, arquivado ou rejeitado?

c. Agrupar deputados por padrão de votação.

Os deputados realmente seguem a orientação partidária ao votar? Existem grupos ideológicos ocultos formados com base no comportamento de voto independentemente do partido?

2. Objetivos do negócio

O objetivo do negócio é desenvolver uma solução inteligente com base em dados abertos da Câmara dos Deputados, que seja capaz de aumentar a transparência, facilitar o monitoramento de atividades parlamentares, apoiar na tomada de decisões tanto por parte dos cidadãos como por parte dos deputados e também ajudar a prever o resultado de votações.

Por meio da aplicação de modelos de Aprendizado de Máquina (Machine Learning), se tem como objetivo:

- a. Mapear o comportamento legislativo, revelando padrões de votações e agrupando deputados por alinhamento ideológico; (Análise Comportamental)
- b. Prever o posicionamento de parlamentares em votações futuras, com base em seu histórico e no conteúdo das propostas; (Análise Comportamental)
- c. Identificar redes de influência e alianças políticas, com base em coautorias e comportamentos de votação similares; (Análise Comportamental)
- d. Antecipar desfecho de projetos de lei, classificando-os como propensos à aprovação, arquivamento ou rejeição; (Previsão)
- e. Detectar gastos parlamentares fora da curva, auxiliando na identificação de possíveis irregularidades; (Detecção de Anomalias)

Ao final do projetos, espera-se entregar uma base analítica sólida e modelos preditivos capazes de gerar relatórios, insights e simulações que demonstrem o potencial do uso de conceitos de Aprendizado de Máquina nos mais diversos campos da sociedade.

3. Descrição do cenário atual ao problema proposto

Nas últimas décadas, a transparência no processo legislativo brasileiro avançou consideravelmente, impulsionada por leis como a Lei de Acesso à Informação (Lei nº 12.527/2011) e por iniciativas de dados abertos. A Câmara dos Deputados disponibiliza, por meio de sua API oficial, uma ampla gama de dados públicos, incluindo informações sobre proposições legislativas, votos, gastos parlamentares e autoria de projetos de lei. Isso torna possível a realização de análises automatizadas e baseadas em dados sobre o comportamento político dos deputados federais.

Apesar dessa abundância de dados, o uso efetivo de Machine Learning para extrair valor ainda é limitado em muitos contextos. Existem iniciativas acadêmicas e jornalísticas que exploram padrões de votação ou destacam gastos atípicos de parlamentares, mas muitas delas utilizam métodos estatísticos tradicionais ou análises descritivas. Isso cria uma oportunidade de aplicar técnicas mais sofisticadas de Machine Learning para identificar padrões, prever comportamentos ou revelar relações ocultas.

Portanto, no cenário político brasileiro, observa-se que é difícil uma análise puramente manual ou baseada em regras fixas. E para os objetivos propostos neste projeto, o uso de Machine Learning se mostra promissor.

4. Hipóteses do projeto

a. Hipóteses para a previsão de aprovação de projetos de lei:

- Treinar modelos coletando informações disponíveis quanto a votação, autor, partido, histórico legislativo, e avaliar seguindo as métricas de **acurácia, precisão e recall**;
- Validar se o modelo consegue manter o mais próximo da performance com dados de diferentes legislaturas;

- Verificar a consistência com o passar do tempo, ou seja, se um modelo criado com dados até 2018 continua performando bem com os dados de 2019 em diante por exemplo.

b. Detecção de gastos anômalos na cota parlamentar:

- Usar algoritmos de detecção de anomalias como o **Isolation Forest**, **Autoencoders**);
- Estabelecer uma baseline com dados “normais” e testar a sensibilidade dos algoritmos em detectar valores fora da curva;
- Verificar casos identificados como anômalos e que são comprovados e comparar com os resultados avaliados da previsão.

c. Agrupamento de deputados por padrão de votação:

- Utilizar técnicas de agrupamento como o **K-Means**, **DBSCAN**, etc.
- Verificar se os grupos formados pelo modelo coincidem com outros aspectos como partidos, votação, etc.
- Analisar se os agrupamentos variam significativamente ao longo do tempo ou durante eventos políticos importantes