

## Projeto de Análise e Previsão de Sorteios

Este projeto é uma plataforma completa de análise de dados e Machine Learning, desenvolvida em R com Shiny, para explorar, entender e prever resultados de sorteios.

### 1. Objetivo do Projeto

O objetivo central é transformar dados brutos de sorteios em insights acionáveis e previsões estatísticas. A ferramenta foi construída para:

- Analisar Tendências: Identificar padrões visuais nos dados históricos.
- Descobrir Relações Ocultas: Utilizar clusterização para encontrar grupos de animais com comportamentos de sorteio semelhantes.
- Prever Resultados: Treinar e comparar múltiplos modelos de Machine Learning para prever a dezena sorteada com a maior acurácia possível.
- Apresentar Resultados: Disponibilizar todas as análises em um dashboard interativo e de fácil utilização.

### 2. Coleta de Dados (Web Scraping)

Para garantir que as análises sejam baseadas em dados reais e atualizados, foi desenvolvido um processo de Web Scraping.

- O que é? É uma técnica automatizada para extrair informações de sites da internet.
- Como foi feito? Um script em Python foi criado para extrair os dados relevantes (animal, dezena, período, etc.).
- Resultado: Todo o dado coletado é salvo no arquivo Pasta1.xlsx, fonte de dados principal do projeto.

### 3. Estrutura e Scripts

train\_models.R

- Carrega e limpa os dados do arquivo Pasta1.xlsx.
- Treina os 6 modelos de Machine Learning (XGBoost, Random Forest, etc.).
- Realiza análise de Clusters.
- Salva resultados em arquivos .rds.

appBicho.R

- Carrega arquivos .rds do train\_models.R.
- Exibe gráficos, tabelas e previsões de forma organizada.

### 4. Seleção e Lógica dos Modelos

Modelos Preditivos:

- XGBoost
- Random Forest
- SVM
- KNN
- Naive Bayes
- Regressão Multinomial

Análise de Clusters:

- Calcula probabilidade sequencial entre animais.
- K-Means agrupa animais com assinaturas semelhantes.

### 5. Resultados e Desempenho

- Melhor Modelo: XGBoost apresentou a maior acurácia.
- Insights: Clusters revelam padrões de proximidade não aleatórios.

## 6. O Aplicativo (Dashboard)

- Visão Geral: Gráficos de frequência.
- Clusters: Gráficos e tabelas interativas.
- Sequência: Previsão de sucessores.
- Comparação de Modelos: Acurácia dos 6 modelos.
- Previsões Futuras: Previsão com novos\_resultados.xlsx.
- Tabela de Animais: Consulta rápida.

## 7. Conclusão

A ferramenta combina scraping, limpeza, Machine Learning e análise de clusters para revelar padrões e apoiar decisões baseadas em dados. A separação entre scripts garante eficiência e rapidez para o usuário final.