**Problema**

"Uma empresa da área de transporte e logística deseja migrar sua frota para carros elétricos com o objetivo de reduzir os custos. Antes de tomar a decisão, a empresa gostaria de prever o consumo de energia de carros elétricos com base em diversos fatores de utilização e características dos veículos."

O conjunto de dados “FEV-data-Excel” inclui carros que, a partir de 2 de dezembro de 2020, poderiam ser adquiridos na Polônia como novos em um revendedor autorizado e aqueles disponíveis em pré-venda pública e geral.

O objetivo é construir um modelo de Machine Learning capaz de prever o consumo de energia de carros elétricos com base em diversos fatores, tais como o tipo e número de motores elétricos do veículo, o peso do veículo, a capacidade de carga, entre outros atributos.

**Metodologia**

Foram feitas transformações e diferentes testes para determinar as variáveis independentes mais significativas para explicar a variável dependente alvo (o consumo), e a partir delas foi possível construir o modelo estatístico.

Para o treinamento dos modelos, é necessário separar os dados entre treino e teste, escolher o melhor modelo pela métrica *Adjusted R-squared* e fazer a aplicação final.

Problemas de regressão, como este, existem quando precisamos prever um valor numérico específico. Usando a linguagem R, foram testados modelos contendo três tipos diferentes de algoritmos para a regressão: Regressão Linear, SVM (*Support Vector Machines*) e GBM (*Gradient Boosting Machine*).

**Resultado**

O modelo 5 (Regressão Linear) foi escolhido por ter tido o melhor desempenho e ser o mais generalizável, podendo ser aplicado no dataset original mesmo para prever os casos em que existiam dados de variáveis faltantes. Este conseguiu precisão de 92% de *Adjusted R-squared* e também serve para prever o consumo dos carros que não tinham dados históricos da variável alvo disponíveis. No dataset “Previsão” está inclusa uma coluna representando o resultado.