Previsão do estado de carga da bateria

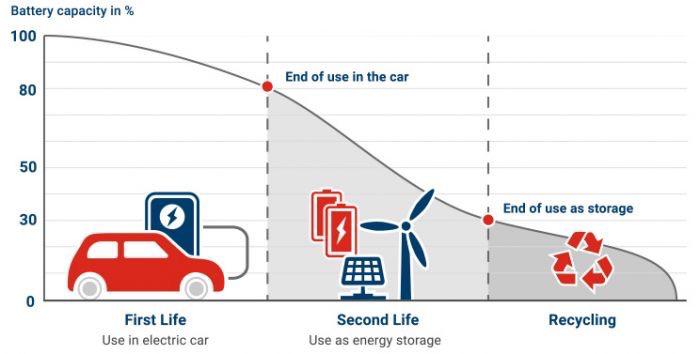
# Contexto

Atualmente, uma das únicas maneiras de avaliar como a vida útil de uma bateria diminui com o passar dos anos é fazendo o ciclo completo de cada célula através de bancadas de teste, o que é muito caro e leva tempo, se tornando uma alternativa menos atrativa.

Através de algoritmos de machine learning, podemos ter uma predição da vida útil da bateria de acordo com as variáveis geradas em viagens teste de um carro elétrico como: tensão, temperatura da bateria, corrente e etc, para assim, analisar de onde vem as maiores degradações da bateria a fim de apontar possíveis melhorias da sua utilização sem perdermos a performance que o carro entrega ao motorista.

Os impactos de prever o ciclo da bateria e gerar dados para trabalhar na extensão da vida útil das baterias, o que não está apenas relacionado a diminuição dos gastos financeiros, mas principalmente, a redução ao impacto ambiental gerado pelo descarte das baterias no meio ambiente.

Esses impactos são mais significativos nas baterias automotivas onde a utilização é mais severa, tendo grandes impactos com a temperatura, quantidade de carga e descarga e a profundidade de descarga, sendo assim para se ter uma autonomia dentro dos parâmetros ideias liberadas pelo fabricante, recomenda-se a utilização da bateria com a capacidade entre 75% à 80% da capacidade nominal da bateria, depois disso a bateria segue para outros tipos de utilização e reciclagem. É nesse contexto que vamos trabalhar e apresentamos nosso problema de negócio.

Figura 1: Utilização da bateria em relação a sua capacidade de carga

# Problema de negócio

Sabemos que as baterias são usadas em quaisquer sistemas eletrônicos, com uma relevância muito maior atualmente em celulares e carros. No contexto atual onde empresas estão buscando créditos de carbono, que é uma moeda em ascensão no Brasil, responsável por reduzir em um milhão de toneladas as emissões de CO2 no país e que, em breve, passará a ser negociada em uma bolsa de valores exclusiva. O ativo sustentável ganha apelo no setor automotivo e da mobilidade, com cada vez mais empresas que apostam na solução como parte da estratégia ESG (sigla em inglês para Governança Ambiental, Social e Corporativa). Nesse contexto, os carros têm ganhado um enfoque maior nos últimos tempos e as pesquisas e popularização dos carros elétricos ou híbridos vem se avançando em larga escala.

As baterias são ricas em como zinco, lítio, níquel, chumbo, mercúrio, cobalto, cádmio, ferro, cobre, manganês e terras-raras, o que causa grandes impactos ambientais no seu descarte. O descarte inadequado faz com que a capsula que protege os metais se danifique, causando assim a liberação no meio ambiente, gerando também líquidos não biodegradáveis e não decomponível gerando uma contaminação quase que permanente no meio ambiente. Essa poluição não se retém apenas ao solo, mas avança a lençóis freáticos e ao ar, através de processos químicos que liberam em forma de gases alguns desses metais como é o caso do mercúrio.

Por esse motivo o ideal seria implementar ao máximo os três “R.s”, Reduzir, Reutilizar e Reciclar, nas baterias automotivas. Existem algumas opções, como reutilizar essas baterias em outros segmentos, como nobreaks, aparelhos hospitalares e celulares. Quando as baterias são retiradas do carro, não quer dizer que elas não prestam, elas só não atendem certos parâmetros porque são muito exigidas, mas continuam sendo boas para utilizar em outros segmentos.

Aplicar um modelo de machine learning vem crescendo amplamente na indústria automotiva a fim de prever valores essenciais que são causadores do envelhecimento precoce da bateria e consequentemente esses estudos trazem benefícios como:

* Prolongar a vida útil da bateria em até 26%
* Otimização do uso de bateria até o fim da vida útil real.
* Reduzir o número de células defeituosas
* Diminuir o descarte (importante por ser um lixo tóxico)

Com o uso de machine learning podemos confiar em um uso real das baterias e determinar se precisa reparar ou substituir peças mais cedo para evitar falhas ou mais tarde preservar custos indesejados.

# Solução

Antes de descrevermos uma solução para o problema, precisamos apresentar algumas definições importantes.

O estado de carga da bateria (representada pela sigla SOC) é um dos parâmetros mais importantes para as baterias, mas sua definição apresenta muitos problemas diferentes, em geral o SoC é definido como a relação entre sua capacidade atual e a capacidade nominal. A capacidade nominal é fornecida pelo fabricante e representa a capacidade máxima de carga que pode ser armazenada na bateria. O SoC pode ser definido da seguinte forma:

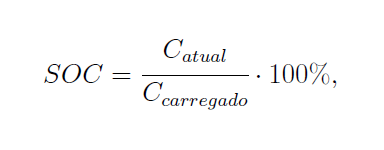


Figura 2: Cálculo de SOC

Estado de carga (SOC) é o nível de carga de uma bateria em relação a sua capacidade, as unidades SOC são pontos percentuais (0% = vazio e 100% = cheio). Onde Catual representa a capacidade da bateria em estado atual e Ccarregado representa a capacidade da bateria em estado carregado (capacidade reversível).

A solução do problema de negócio será apresentada na forma de uma regressão utilizando a base de dados de viagens feitas por um carro elétrico. O modelo gerará dados tanto para estudarmos o comportamento da bateria e prever os parâmetros que influenciam no estado de carga da bateria, quanto poderá ser utilizado na criação de outros modelos que podem atuar no carro criando diretrizes que farão com que o carro possa, sem perder a sua funcionalidade e segurança ao motorista, tomar ações que impeçam a degradação da bateria.

# A base de dados

Nossa base de dados escolhida foi um histórico de viagens e captação de variáveis do carro elétrico BMW i3 com as seguintes características de bateria proposta pelo fabricante:

Tabela

Descrição gerada automaticamente

Figura 3: ados do fabricante sobre a bateria utilizada no veículo

Além disso com o veículo instrumentado nos testes de estrada, levantou-se uma série de parâmetros medidos e armazenados no dataset que iremos trabalhar e suas respectivas unidades de medidas:

|  |  |
| --- | --- |
| Time | Tempo em Segundos |
| Velocity | Velocidade em Km/h |
| Elevation | Altitude em metros |
| Throttle | Posição do pedal do acelerador em % |
| Motor Torque | Torque do Motor em N.m |
| Longitudinal Acceleration | Aceleração longitudinal m/s² |
| Regenerative Brake | Sinal do freio regenerativo |
| Battery Voltage | Tensão da Bateria em V |
| Battery Current | Corrente da bateria em A |
| Battery Temperature | Temperatura da bateria em ºC |
| Max battery Temperature | Temperatura máxima da bateria em ºC |
| Soc – State of Charge | Estado da bateria em % |
| Displayerd SoC | Estado de carga mostrado em % |
| Min SoC | Mínimo estado de carga em % |
| Max SoC | Máximo estado de carga em % |
| Healting Power CAN | Potência do aquecedor na rede CAN Kw |
| Healting Power LIN | Potência do aquecedor da rede LIN kw |
| Requested Heating Power | Potência registrada em W |
| AirCon Power | Potência do ar-condicionado em Kw |
| Heater Signal | Sinal do aquecedor binário |
| Heater Current | Corrente do Aquecedor A |
| Ambient Temperature | Temperatura ambiente em ºC |
| Coolant Temperature Heatercore | Temperatura do líquido do radiador de cabine em ºC |
| Requested Coolante Temperature | Temperatura solicitada do líquido de arrefecimento em ºC |
| Coolant Temperature Inlet | Temperatura do líquido de arrefecimento na entrada ºC |
| Heat exchanger Temperature | Temperatura do trocador de calor em ºC |
| Cabin Temperature Sensor | Temperatura do sensor da cabine em ºC |

# EDA – Análise exploratória dos dados

A nossa base de dados apresenta 31645 observações com 23 variáveis, dentre os quais não encontramos nenhum nulo.

Após a análise dos dados, fizemos uma análise de correlação para verificar o peso de cada variável na predição do SOC entre todas as variáveis e para termos uma melhor seleção das features que mais se ajustam no modelo de regressão, além disso, utilizamos biblioteca Random Forest utilizando a “feature importance” para avaliarmos quais os pesos dentre todas as variáveis e concluímos que as variáveis de temperatura, tensão e velocidade foram as que mais tiveram relevância para utilizarmos na regressão, conforme gráfico abaixo:

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Gráfico

Descrição gerada automaticamente

Fazendo uma análise da distribuição dos dados em função do estado de carga vimos que algumas variáveis têm um impacto positivo e outras tem um impacto negativo, a fim de darmos uma explicação geral dentre todas as variáveis podemos concluir que o estado de carga da bateria (SOC) degrada em função do tempo e em função da demanda de utilização, ou seja, quanto maior a demanda de potência, tensão e carga, mais rapidamente o SOC vai cair para 0%.

Interface gráfica do usuário, Aplicativo, Linha do tempo

Descrição gerada automaticamente com confiança média

Gráfico

Descrição gerada automaticamente

Gráfico, Histograma

Descrição gerada automaticamente

Plotando a distribuição da variável target, podemos ver que a distribuição é uma distribuição bimodal onde tenho dois grandes picos, esses dois grandes picos podem ser uma característica de que o veículo esteja utilizando o freio regenerativo o que, como consequência gerará uma carga extra na bateria e consequentemente apresentando duas modas. Outro ponto que vale ressaltar é que não temos uma curva normal, porém, existem estudos que curva não simétricas tem um bom comportamento em resultados estatísticos e muitas vezes modificar a variável target não é a solução, nesse caso manteremos a distribuição não normal.

Após o levantamento e análise das curvas, fizemos um pré-teste utilizando o método dos mínimos quadrados do StatsModels e analisando estatisticamente as variáveis escolhidas vemos que meus resultados estão abaixo dos 5% da população, ou seja, não podemos rejeitar a hipótese nula H0, enquadrando todas essas variáveis como tendo impacto significativo na SoC, que é a nossa variável target.

Interface gráfica do usuário, Tabela

Descrição gerada automaticamente com confiança média

Analisando os resíduos em função dos valores estimados, podemos observar a independência e a homoscedasticidade, que os resíduos se distribuem de maneira razoavelmente aleatória e com a mesma amplitude do zero.

E analisando o R^2 de 0.983 e o ajustado também de 0.983, podemos afirmar que as variáveis explicativas se ajustam a linha da regressão.

Gráfico, Histograma

Descrição gerada automaticamenteGráfico, Gráfico de dispersão

Descrição gerada automaticamente

# Preparando o modelo para regressão

Para iniciarmos a regressão, utilizamos o dataset com 31645 observações onde 30% foram separados para o teste do modelo dentre os modelos escolhido, onde fizemos uma análise de regressão utilizando os modelos abaixo que daremos uma breve explanação

## ElasticNet

Esse modelo tem um compromisso entre a regressão Ridge e Lasso, combinando os dois tipos de Penalties. Os coeficientes elastic net minimizam a qualidade.

Desenho de personagem

Descrição gerada automaticamente com confiança baixa

A vantagem deste modelo é permitir uma efetiva regularização com penalidade do tipo Ridge, especialmente em grupos de preditores correlacionados, aliados à qualidade da seleção de características através do Lasso. A figura abaixo ilustra esta penalização, comparando-a com ridge e lasso.

Gráfico, Gráfico de radar

Descrição gerada automaticamente

## Ridge e Lasso

A regressão Ridge falha na parcimônia do modelo, pois ainda que as estimativas dos parâmetros sejam muito próximas de zero, elas nunca atingem exatamente esse valor. Assim, todos os preditores, ainda que com pouco peso, permanecem no modelo.

A regressão Lasso é uma alternativa que contorna essa desvantagem, os coeficientes lasso, β^Lassoλβ^λLasso, minimizam a quantidade.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente com confiança média

Embora possa parecer uma pequena modificação, as implicações práticas são significantes. Neste caso, dependendo do valor de λλ, algumas estimativas dos parâmetros são realmente 0. Assim, o penalty ℓ1ℓ1 funciona também como um selecionador de vereáveis. A figura abaixo ilustra essa situação, a estimativa de β1β1 via penalização Lasso é exatamente igual a 0, enquanto via Ridge nunca atingirá esse valor.

Gráfico de radar

Descrição gerada automaticamente com confiança baixaFigura 4: Ridge e Lasso respectivamente

## Regressão SVR

A ideia básica do SVR é mapear um conjunto de dados explicativos em um espaço multidimensional através de um mapeamento não-linear (usualmente utilizando kernels) e então realizar uma regressão linear neste espaço multidimensional transformado.

O SVR é extremamente robusto, mesmo em espaços com muitas dimensões, isso se da devido a optimização não depender da dimensão do espaço de input. O SVR depende apenas de um subconjunto dos dados de treino, pois a função custo para a construção do modelo ignora os dados de treino perto da previsão do modelo.

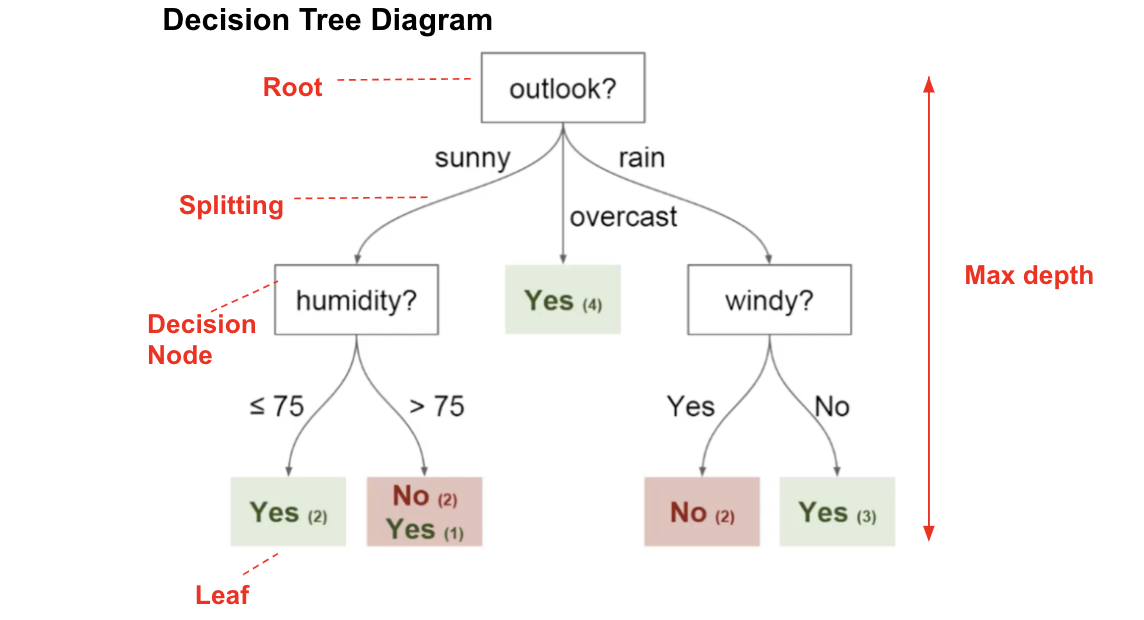
SVR sobre o modelo de regressão linear usual tem a vantagem de utilizar uma grande variedade de funções que se adéqua aos diferentes modelos, fornecendo assim um bom ajuste aos diversos tipos de ruídos que os dados possam apresentar.

## Tensorflow

Tenserflow é uma biblioteca de código aberto feita pela google e apresenta uma solução capaz de preparar e criar os modelos muito populares para aplicativos de aprendizado de máquina, especialmente deep learning. A principal característica positiva para o modelo é a capacidade dele se comunicar com mais de uma linguagem de programação, isso faz com que a biblioteca se torne versátil e pode ser usada para se comunicar com outras linguagens de programação.

## Random Forest

É um método de aprendizado de máquina para problemas que envolvem classificação ou regressão. Ele se baseia em uma coleção de árvores de decisão, podemos pensar como uma representação gráfica para um determinado processo de decisão. Essas árvores são formadas por nós que armazenam informações. O nó raiz que possui o nível hierárquico e se ramifica nos nós filhos, como representado na imagem abaixo.



# Resultados

Aplicamos cada uma dessas regressões nos nossos dados e avaliamos cada um igualmente para encontrar o que melhor atende nosso problema, para isso usamos algumas métricas para avaliar nosso modelo regressão escolhido, das quais vamos focar nas seguintes métricas: RMSE (Raiz quadrada do erro quadrático médio), MAE (Erro absoluto médio) e o R^2 (R-quadrado).

Tabela

Descrição gerada automaticamente

# Conclusão e próximos passos

O modelo que apresentou maior robustez na predição foi o Random Forest, tendo em vista que foi o modelo que obteve o melhor métrico MAE igual a zero (a escolha do MAE foi devido a uma punição menos sevara a possíveis outliers, em caso de falhas nos sensores).

Como o modelo de regressão se mostrou eficaz podemos dizer que a aplicação do método de regressão escolhido para reduzir o impacto de degradação na vida da bateria poderá ser aplicado no módulo da bateria a fim de controlar parâmetros que possam influenciar negativamente numa queda brusca do SOC, sendo assim predições fora dos parâmetros ideais podem ter intervenções instantâneas através de programações específicas, ajustando assim fatores impactantes em tempo real.

* Ribeiro, João Gabriel Rodrigues; Santos, Matheus Felipe e Chagas, Nicholas Silva. O Impacto ao meio ambiente pelo descarte incorreto de pilhas e baterias. Pouso Alegre;
* Murnane, Martin e Ghazel, Adel. A Closer Look at State of Charge (SOC) and State of Health (SOH) Estimation Techniques for Batteries;
* Hyung-Kyu Kim, Byung-Uk Do e Il-Gue Kim, Battery recycling: Carmakers, batery companies in new turf war, 2021, <https://www.kedglobal.com/battery-recycling/newsView/ked202108160002> , 07 de setembro de 2022;
* Fernandez, Bruno Oziel. Predição do estado de saúde de baterias de Íon-Lítio por meio de técnicas de aprendizagem de máquina e séries temporais. Foz do Iguaçu, 2021;
* [http://cursos.leg.ufpr.br/ML4all/apoio/Regularizacao.html#3)\_penaliza%C3%A7%C3%A3o\_elastic\_net](http://cursos.leg.ufpr.br/ML4all/apoio/Regularizacao.html#3)_penalização_elastic_net)
* https://www.vebuso.com/2020/01/decision-tree-intuition-from-concept-to-application
* <https://lamfo-unb.github.io/2020/07/08/Random-Forest/>
* <https://automotivebusiness.com.br/pt/posts/mobility-now/empresas-automotivas-compram-creditos-de-carbono-como-parte-da-estrategia-esg/>
* https://autoesporte.globo.com/um-so-planeta/noticia/2021/04/da-para-reciclar-as-baterias-de-ions-de-litio-que-equipam-os-motores-de-carros-eletricos.ghtml