

Universidade Presbiteriana Mackenzie

Faculdade de Computação e Informática - Ciência de Dados

Projeto Aplicado II – Análise do comportamento de defeitos apresentados por alternadores
automotivos em período de garantia

Membros do grupo: Carlos Oliveira, Felipe Ferraz, Erick Isidoro

São Paulo, 05 de outubro de 2024

Sumário

Glossário	3
Introdução	4
1. Objetivo.....	6
1.1 - Objetivo do estudo.....	6
2. Empresa, contexto e problema do estudo	7
2.1 - Empresa.....	7
2.2 - Contexto e problema do estudo.....	8
3. Metadados.....	10
3.1 - Referências de aquisição	10
3.2 - Descrição da origem	10
3.3 - Descrição	10
4. Análise dos Dados.....	12
4.1 - Linguagem que será utilizada no projeto	12
4.2 - Verificação de relação em R	12
4.3 - Analise tratamento e treinamento em Python	14
5. Conclusão.....	16
6. GitHub.....	16
6.1 - Link para acesso ao projeto no GitHub.....	16
Tabela de imagens.....	17
Referências Bibliográficas	18

Glossário

Alternador: Peça responsável por converter energia mecânica em energia elétrica, o alternador é fundamental para o funcionamento adequado do veículo. O alternador automotivo também tem a função de carregar a bateria automotiva.

Dataset: Coleção de dados textos e numéricos normalmente estruturados, sobre determinado assunto.

MIS: Months in service , meses em serviço; É o tempo que demora em meses para uma peça falhar a partir da venda do veículo.

Introdução

Neste projeto, foi proposto o desafio de aprimorar a diagnose de pós-venda de uma empresa do setor automotivo, especializada na fabricação de alternadores. A empresa busca melhorar sua capacidade de prever o tempo médio até a ocorrência de falhas em seus produtos, a fim de identificar a necessidade de intervenções preventivas que possam reduzir a incidência de reclamações. Com isso, espera-se não apenas fortalecer a imagem da empresa perante os clientes, mas também otimizar os custos operacionais.

A análise será realizada com dados reais, visando aumentar a precisão das predições e agregar valor prático ao projeto, resultando em uma solução que contribua para a melhoria da performance dos alternadores no mercado.

**Projeto Aplicado II - Análise do comportamento de defeitos apresentados por
alternadores automotivos em período de garantia**

1. Objetivo

1.1 - Objetivo do estudo

Neste projeto queremos realizar uma análise descritiva e preditiva dos dados coletados de peças com defeito a fim de auxiliar a resolução para problemas de diagnose relacionados a pós-venda ou também chamada garantia, quando uma empresa vende um produto e concede um período de garantia, precisando acompanhar o desempenho do mesmo em campo.

O projeto tem a missão de esclarecer por meio de análises pontos importantes, e criar um modelo de treinamento para obter uma visão do futuro utilizando machine learning.

2. Empresa, contexto e problema do estudo

2.1 - Empresa

A Valeo é uma indústria de peças automotivas, fornecedora de produtos originais e parceira de todas as montadoras. Sua história começou em 1923 em Paris e hoje a empresa atua em 29 países, se consolidando como a 5ª maior fabricante de peças do Mundo.

No Brasil, sua história começou em 1974 e hoje tem 12 sites de produção, 1 centro de distribuição dedicado para o mercado de reposição (Valeo Service) e 3 centros de pesquisa e desenvolvimento de produtos (P&D). Suas fábricas estão situadas em 8 cidades diferentes: São Paulo, Mauá, São Bernardo do Campo, Campinas, Itatiba, Gravataí, Cachoeirinha e Caxias, empregando mais de 3.000 funcionários.

A Valeo desenvolve soluções e produtos de alta tecnologia que colocam o Grupo no epicentro das três revoluções que afetam a indústria automotiva atual: eletrificação, veículos autônomos e mobilidade digital. À exemplo disso, tem-se como destaque o desenvolvimento de soluções para carros levemente híbridos.

Além dos alternadores, objeto desse estudo, são fabricados ainda: radiadores, eletroventiladores, intercoolers, ar-condicionado, compressores, evaporadores, faróis, lanternas, palhetas, atuadores e embreagens.

A missão da empresa é ser “fornecedora automotiva e parceira de montadoras em todo o mundo. Como empresa de tecnologia, concebemos soluções inovadoras para a mobilidade inteligente, com especial enfoque na condução intuitiva e na redução das emissões de CO₂. O Grupo também fornece e distribui peças de reposição para montadoras e operadores independentes de reposição”.

2.2 - Contexto e problema do estudo

A fabricante escolhida pelo grupo fornece peças automotivas para diferentes montadoras, que por sua vez as utilizam na fabricação de seus diversos modelos de veículos. Esses automóveis são montados em diferentes plantas e vendidos no mercado brasileiro e no exterior.

Atualmente, há a necessidade por parte da Valeo de analisar o comportamento de defeitos apresentados em suas peças, na tentativa de identificar possíveis padrões, correlações e causas, e assim tomar decisões a fim de melhorar seus processos e aumentar a qualidade de seu produto.

A empresa conta com uma área de Qualidade, mas não é uma área com iniciativas de Data Science, sendo assim, carece de representação gráfica da performance das peças em campo, demonstração da evolução dos dados, tempo hábil para obtenção dos dados e outras questões de análise descritiva, diagnóstica, preditiva e prescritiva.

Utilizando as técnicas de pensamento computacional, dentre as diferentes peças produzidas pela empresa, o grupo escolheu trabalhar apenas com os alternadores, pois seria difícil fazer uma análise com diferentes peças, dessa forma quebrou-se um problema maior em um menor.

A identificação de um problema no alternador segue o padrão de ser realizada em uma concessionária autorizada no Brasil ou no exterior, a partir desse momento são gerados os dados que constam no metadados provenientes do reparo. A garantia da peça em questão também segue o padrão de passar a vigorar a partir da compra do veículo, ou seja, ela independe

de sua data de fabricação. Essas informações são importantes para garantir que os dados partam de um mesmo protocolo, garantindo sua confiabilidade.

A identificação de um problema no alternador segue um padrão estabelecido: o cliente percebe o defeito, leva o veículo à concessionária, onde é realizada uma análise. Caso necessário, o reparo é efetuado, e, com o problema resolvido, o carro é devolvido ao cliente.

Durante esse processo, são coletados dados que são disponibilizados tanto à montadora quanto ao fabricante das peças — neste caso, o alternador. Se a montadora desejar uma análise mais detalhada, a peça é resgatada e enviada ao fabricante para uma avaliação mais aprofundada.

É responsabilidade do fabricante analisar as falhas e propor melhorias, com o objetivo de reduzir reclamações e aumentar a vida útil da peça.

Nosso objetivo é desenvolver um modelo preditivo e descritivo que auxilie o fabricante a avaliar a eficácia de suas ações e verificar se a média dos tempos até falhas (MTBF) está aumentando ou diminuindo.

3. Metadados

3.1 - Referências de aquisição

O dataset utilizado está no formato “.csv”, vem de um banco de dados nomeado “Wings” e é privado, não sendo possível acesso ao público. Um dos membros do grupo, Felipe Ferraz, trabalha na empresa, é usuário desse banco de dados e obteve autorização para disponibilizar o dataset com as restrições descritas a seguir.

Os dados levantados compreendem o período de fevereiro de 2023 até 22 de fevereiro de 2024 e não possuem validade. Por questões de segurança, sigilo, sensibilidade e outras previstas na LGPD, os nomes das montadoras, seus modelos e plantas fabris foram anonimizados mantendo as correspondências originais.

Parte da base de dados em questão é alimentada por pessoas, e alguns dados apresentaram inconsistência, portanto, foram devidamente tratados.

3.2 - Descrição da origem

Quando o cliente leva seu veículo na concessionária, e esta identifica o defeito na peça, ela é trocada e a própria autorizada faz a descrição no sistema da respectiva montadora, alimentado com os dados necessários a ocorrência. Esses dados são cruzados com os do sistema da Valeo e dão origem a base de dados, tal como a utilizada nesse projeto.

3.3 - Descrição

O dataset possui 3111 linhas com 21 colunas, e nele estão registrados os defeitos apresentados pelos alternadores de acordo com as seguintes variáveis organizadas pelo seu tipo:

- VARCHAR: cliente (customer), planta do cliente (customer plant), modelo do veículo (vehicle name), família do alternador (sub family), sobressalente ou original (spare or original), nr motor (engine nr), país do reparo (repair country), clima (climate);

- INT: pn cliente (customer pn), quilometragem (kms), fator técnico (technical factor), tempo em serviço da peça em meses (mis) e km/ano (km/year);
- DATE: data de produção da peça (production date), início da garantia da peça (inservice date), data do reparo (repair date), data de gravação do cliente (customer recording date), data de carga (upload date);
- FLOAT: custo total de garantia (total warranty cost €), custo de mão de obra (labor cost €), custo de material (material cost €);

4. Análise dos Dados

4.1 - Linguagem que será utilizada no projeto

Estaremos desenvolvendo a análise do projeto a linguagem em R e Python, ferramentas muito importantes para análise estatística, descritiva, preditiva, aprendizado de máquina e outras ferramentas gráficas e matemáticas muito importantes.

Em python utilizaremos as bibliotecas, Pandas – ferramenta importante na leitura dos dados , agrupamento, seleção, agregação, facilita análise estatística, Numpy – biblioteca poderosa para matriz dimensional, operações matemáticas, Matplotlib e Seaborn – bibliotecas importantes para criação gráfica estatística e customização, Request – biblioteca importante para requisições, autenticações, uploads e downloads, MinMaxScaler – ferramenta pré-processamento em machine learning, train_test_split – ferramenta que divide dataset em conjunto de treino , KNeighborsClassifier – ferramenta com métodos de Kvizinhos, classification_report – ferramenta que fornece análise detalhada de desempenho de um modelo de classificação com F-1, recall e outros.

4.2 - Verificação de relação em R

Foi realizada a análise em R para verificação de relação entre os dados Customer e KM/Year, conforme resultados que demonstraremos abaixo, identificamos que há uma diferença estatisticamente significativa entre os grupos de Customer em relação KM/Year, podendo rejeitar a hipótese nula, pois há uma associação Significativa entre Customer e KM/Year, resultados que esperávamos já que existe nos grupos uma grande diferença de registros para KM/Year.

Para análise tivemos que remover dois valores nulos, abaixo imagem dos resultados obtidos.

DataTeste

	Customer	KM.Year
1	Make A	7082
2	Make A	9709
3	Make A	16191
4	Make A	17823
5	Make A	14858
6	Make A	14190
7	Make A	5452
8	Make A	9691
9	Make A	2424

Showing 1 to 10 of 3,108 entries, 2 total columns

```
R 4.3.2 - C:/Users/Carlos/Desktop/Aulas R/
> setwd("C:/Users/Carlos/Desktop/Aulas R")
> DataTeste <- read.csv("Dados_alternadores_garantia2.csv", header = TRUE, sep = ";", colClasses = c('factor', 'numeric'))
> str(DataTeste)
'data.frame': 3108 obs. of 2 variables:
 $ Customer: Factor w/ 4 levels "Make A","Make B",...: 1 1 1 1 1 1 1 1 1 ...
 $ KM.Year : num 7082 9709 16191 17823 14858 ...
> View(DataTeste)
> View(DataTeste)
> One.way <- aov(KM.Year~Customer, DataTeste)
> summary(One.way)

> summary(One.way)
          Df Sum Sq Mean Sq F value    Pr(>F)
Customer    3 1.498e+10  4.993e+09    22 4.27e-14 ***
Residuals 3104 7.044e+11  2.269e+08
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Name	Type	Value
One.way	list [13] (S3: aov, lm)	List of length 13
coefficients	double [4]	20094 188 8929 -1360
residuals	double [3108]	-13012 -10385 -3903 -2271 -5236 -5904 ...
effects	double [3108]	-1078222 13314 121212 -10505 -4729 -5397 ...
rank	integer [1]	4
fitted.values	double [3108]	20094 20094 20094 20094 20094 20094 ...
assign	integer [4]	0 1 1 1
qr	list [5] (S3: qr)	List of length 5
df.residual	integer [1]	3104
contrasts	list [1]	List of length 1
xlevels	list [1]	List of length 1
call	language	aov(formula = KM.Year ~ Customer, data = DataTeste)
terms	formula	KM.Year ~ Customer
model	list [3108 x 2] (S3: data.frame)	A data.frame with 3108 rows and 2 columns

4.3 - Analise tratamento e treinamento em Phyton

Para analise e treinamento dos dados em phyton, realizamos o tratamentos e seleção dos dados Customer o qual registra a unidade fabricante, e outros atributos numéricos que pudessem basear o treinamento da máquina, na coluna KM/Year tivemos que transformar os dados em Int64, não foi necessário a remoção de dados pois não tínhamos dados nulos, o treinamento foi realizado em dois modelos para verificar a acurácia, com k-vizinho (3) obtivemos resultado 86%, com o k-vizinho (11) obtivemos um resultado melhor de 89 % de acurácia. (a análise pode ser visualizada completamente no link GitHub no ponto 5.1 deste projeto.

Para os dois modelos aplicamos uma predição para novos 5000 registros e obtivemos resultados bem diferentes.

Para K vizinho 3 obtivemos os resultados.

```
0.8629550321199143
```

Classification Report:

	precision	recall	f1-score	support
Make A	0.12	0.11	0.12	18
Make B	0.67	0.38	0.48	58
Make C	0.22	0.09	0.13	44
Make D	0.90	0.96	0.93	814
accuracy			0.86	934
macro avg	0.48	0.38	0.41	934
weighted avg	0.84	0.86	0.85	934

```
warnings.warn(  
Make D    2523  
Make B    2477  
Name: count, dtype: int64
```

Para K-vizinhos 11 obtivemos os resultados.

```
0.8918629550321199
```

Classification Report:

	precision	recall	f1-score	support
Make A	0.00	0.00	0.00	18
Make B	1.00	0.31	0.47	58
Make C	1.00	0.02	0.04	44
Make D	0.89	1.00	0.94	814
accuracy			0.89	934
macro avg	0.72	0.33	0.36	934
weighted avg	0.88	0.89	0.85	934

```
warnings.warn(  
Make D    5000  
Name: count, dtype: int64
```

Ressaltando a importância de tratamos os problemas do fabricante Make D, levando em conta que o fornecedor não tenha alterado nada em sua forma de fabricar os produtos, os modelos reforçam a continuação dos altos registros em Make D.

5. Conclusão

6. GitHub

6.1 - Link para acesso ao projeto no GitHub

Abaixo tem-se o link do GitHub onde estão compartilhados, dataset, cronograma e outros documentos atualizados do projeto:

Demonstração da imagem do cronograma em tabelas de imagem (Imagem 1 – Cronograma de Atividade), arquivo disponibilizado em link Github.

<https://github.com/Ferraz0Felipe/PROJ-APLICADO-2>

Imagem 1 – Cronograma de Atividade

[illegible]

Referências Bibliográficas

<https://www.valeo.com/pt-br/brasil/> acessado em 26/03/2024.