



Desenvolvimento de um modelo preditivo para sinistros de automóveis baseado em *Naive Bayes*: Estudo de caso numa seguradora do município do Rio de Janeiro

Alex Macedo Teles Silva¹; Guilherme Portilho Joaquim²; Enderson Luiz Pereira Júnior³; Bruno Andrade Martins⁴; Marcos dos Santos⁵.

Resumo: O objetivo deste trabalho é estudar os benefícios na utilização de algoritmos de *Machine Learning* na modelagem de um problema de direcionamento de sinistro de automóveis para as Oficinas Referenciadas na rede da seguradora. O trabalho aborda conceitos de aprendizagem de máquina, aprofundando-se no entendimento dos dados e interpretação do algoritmo supervisionado de *Naive Bayes* para modelagem do problema. Com isso, a análise das variáveis na base de dados dos sinistros foram importante para entender um pouco melhor o perfil dos clientes em escolher a oficina onde pretendem reparar seus automóveis após um sinistro. O modelo apresenta grande margem para otimização, pois obteve bom desempenho quanto a sua acurácia de 88%, bem como outros indicadores de desempenho. Com isso, a princípio, o modelo proposto é apresentado de forma promissora num campo ainda pouco explorado no mercado de seguros.

Palavras-chave: Seguros, *Machine Learning*, *Naive Bayes*, Modelagem

Abstract: The objective of this work is to study the benefits of using Machine Learning algorithms in the modeling of a car loss routing problem for referenced workshops in the insurance company's network. The work addresses machine learning concepts, deepening the understanding of data and interpretation of the Naive Bayes supervised algorithm for modeling the problem. Thus, an analysis of the variables in the claims database was important to understand a little better the profile of customers in choosing a workshop where they intend to repair their cars after a claim. The model has great scope for optimization, as it performed well with its 88% accuracy, as well as other performance indicators. Thus, at first, the proposed model is proposed in a promising way in a field that is still little explored in the insurance market.

Keywords: Insurance, Machine Learning, Naive Bayes, Modeling.

¹ Universidade Estácio de Sá (UNESA), alexmacedo45@gmail.com

² Fundação Instituto de Administração (FIA), guilherme.portilho@outlook.com

³ Universidade do Sul de Santa Catarina (UNISUL), endersonjr@gmail.com

⁴ Secretaria de Estado de Polícia Militar (SEPM), bruno.martins@pmerj.rj.gov.br

⁵ Instituto Militar de Engenharia (IME), marcosdossantos@ime.eb.br

1 Introdução

A revista Apólice (2020), descreve a história recente da humanidade, que está atravessando um momento muito delicado em todos os sentidos devido ao COVID-19. Sendo assim, o cenário que tem se mostrado, é um enorme desafio para todas as formas de organização política, social e econômica. Acredita-se que em qualquer crise também é possível buscar novas oportunidades. Não seria diferente para a indústria do setor securitário, onde novos produtos e serviços devem surgir a partir de novas necessidades.

Os dados atuais do mercado segurador, mostram que as seguradoras e os parceiros corretores têm demonstrado resiliência, mesmo em um período de dificuldades. O faturamento do mercado de seguros brasileiro totalizou R\$ 81 bilhões no primeiro quadrimestre deste ano, revelando alta de 4,9% em comparação ao mesmo período do ano passado, reforçando a solidez do segmento. É extremamente importante ressaltar que também precisou se reinventar do dia para a noite, assim como a grande maioria da população mundial.

No atual momento em que a sociedade está vivendo, diante da necessidade de responder aos desafios gerados pela pandemia mundial ocasionada pelo COVID-19, as seguradoras precisam manter suas operações, mesmo com a severa interrupção por conta da mudança de contexto. Sendo assim, agora devido ao cenário cada vez mais complexo, os gestores devem tomar decisões relevantes com base na análise de dados, tentando mitigar ao máximo se basear em teorias empíricas como diagnóstico para um eventual plano de ação.

O avanço da tecnologia, certamente é uma arma para melhorar problemas empresariais. Pode-se destacar o surgimento da quarta Revolução Industrial, também denominada como Indústria 4.0, onde novas tecnologias habilitadoras, tais como: Internet das Coisas (*IoT*), Computação em Nuvem e *Big Data Analytics* tem ampliado as possibilidades para a utilização de dados digitais provenientes do chão de fábrica na tomada de decisão e otimização dos processos organizacionais.

Neste cenário, um conjunto de ferramentas e técnicas específicas para a coleta, tratamento e análise de grandes volumes de dados torna-se uma alternativa diante das tradicionais abordagens voltadas à excelência operacional.

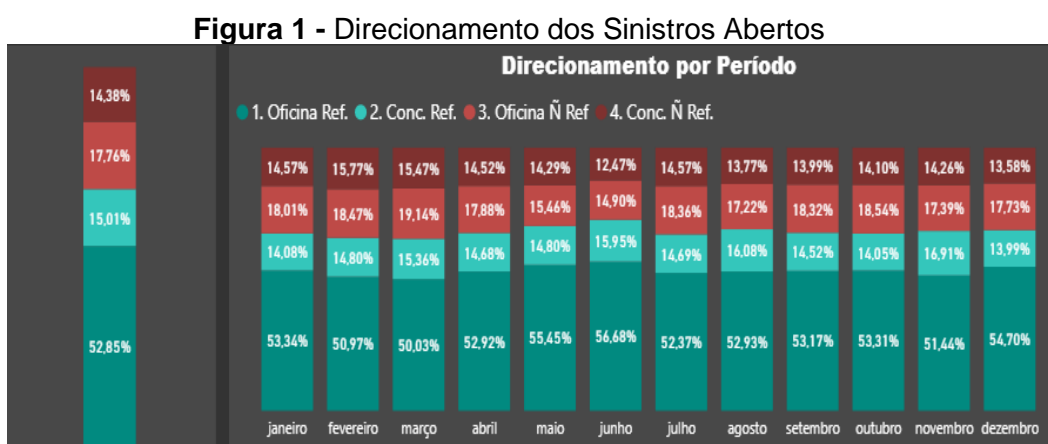
Pode-se destacar a análise de dados preditiva como uma proposta para melhorar a excelência operacional. O mais popular dos tipos de análise de dados é justamente o modelo preditivo, pois como o próprio nome diz, sua essência está na previsão de cenários futuros com base na análise de padrões revelados pela base de dados. Diante disto, é importante saber que, em uma análise preditiva, não é possível prever o que vai acontecer, mas sim, o que deveria acontecer, caso determinadas condições sejam satisfeitas.

Ao decorrer do presente estudo, será construída uma aplicação de modelo classificador de dados supervisionados de *Machine Learning* usando o conceito de *Naive Bayes*. A análise dos dados descritivos será gerada a partir dos dados coletados através do painel gerencial do setor de Gestão da Rede de Oficinas de uma seguradora.

1.1 Entendimento do Problema

A diretoria de operações da seguradora lança mão de alguns indicadores visando medir o desempenho de sua operação visando obter lucros cada vez mais expressivos. A prestação do serviço deve ter cada vez mais qualidade para sua carteira de segurados. Sendo assim, um dos indicadores relevantes do Setor de Rede de Oficinas é o direcionamento dos sinistros. O indicador consiste em saber quantos sinistros abertos do total foram direcionados para Rede Referenciada de Oficinas.

Após uma análise sobre o painel gerencial, foi observado que o indicador de direcionamento de sinistro para Rede de Oficinas Referenciadas está praticamente no mesmo patamar desde o momento em que foi iniciada a análise. Como pode ser observado na Figura 1.



Fonte: Autores, 2021

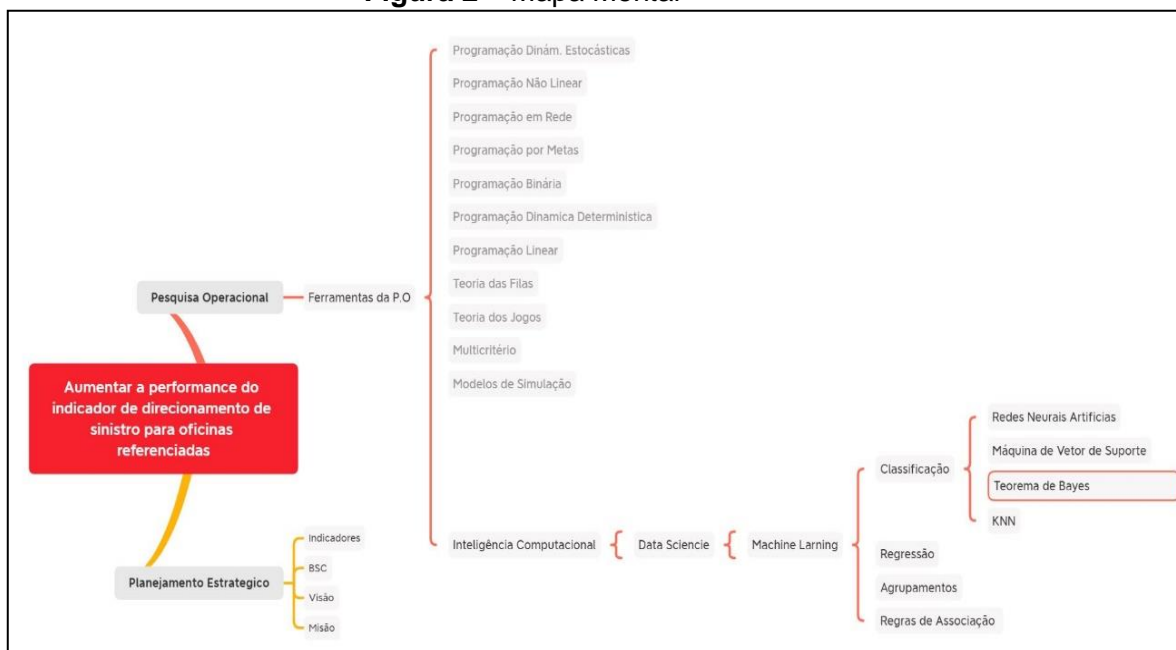
O melhor cenário possível na regra de negócio é que todos os segurados e terceiros sejam assistidos em uma Oficina Referenciada. Vale ressaltar que o cliente tem perante a Lei o direito de escolher onde o seu automóvel deve ser reparado.

O problema do presente trabalho foi elaborado a partir da seguinte pergunta: De qual forma é possível melhorar a performance do indicador de desempenho do direcionamento de sinistro para a Rede de Oficinas Referenciada?

Com isso, se a empresa não conseguir equacionar o problema destacado acima, ou seja, entender o motivo de alguns segurados ou terceiros direcionarem seus veículos sinistrados para uma Oficina Não Referenciada, a empresa diminuirá sua participação de mercado, devido ao alto índice de custo na reparação de cada sinistro e a qualidade do serviço prestado, impactando diretamente na receita da empresa.

Para elucidar melhor a estrutura da pesquisa, foi construído um mapa mental, onde foi elaborado um diagrama para representar ideias, tarefas ou outros conceitos que se encontram relacionados com uma palavra-chave ou uma ideia central, conforme apresentado na Figura 2.

Figura 2 – Mapa Mental



Fonte: Autores, 2021

Tendo como base o Mapa mental, podemos observar o passo a passo de como será apresentado o presente estudo. Para equacionar o problema, foi adotado o método de *Machine Learning* usando o teorema de *Bayes*.

Sendo assim, esta pesquisa se desdobra nos seguintes objetivos:

- Aplicar um modelo preditivo na abertura de um chamado de sinistro, com o objetivo de identificar a probabilidade do segurado ou terceiro escolher uma Oficina Referenciada. A estratégia da diretoria é conceder desconto na franquia com intuito de persuadir o cliente que tenha a probabilidade maior em não levar para Rede de Oficinas Referenciadas, entretanto, para o cliente que tem a probabilidade maior em levar para uma Oficina Referenciada, não será aplicado desconto da franquia.
- Avaliar a performance do modelo aplicado.

A justificativa central do trabalho será aplicada em um modelo de classificação de aprendizagem de máquina com o intuito de aumentar a performance do indicador de direcionamento de sinistros, pois, com o aumento, a empresa obtém uma economia significativa. Um outro ponto é sobre a qualidade dos serviços prestados ao consumidor, pois, quando a reparação do automóvel é feita em uma Oficina Referenciada, a oficina tem um alto nível de cobrança na prestação do serviço por ser uma Oficina Referenciada pela rede. Sendo assim, a seguradora tem um controle com maior amplitude na qualidade prestada.

1.2 Metodologia

Em consonância com Marconi e Lakatos (2016), a pesquisa desenvolvida possui caráter explanatório, visto que tem como principal finalidade desenvolver, esclarecer e modificar conceitos e ideias para a formulação de abordagens posteriores.

Segundo Oliveira (2011), como são feitas validações de hipóteses por meio de uma grande massa de dados estruturados, a pesquisa pode ser caracterizada como quantitativa. Quanto ao método, é classificado como indutivo. Como descreve Freitas (2013), busca-se utilizar uma particularidade e então generalizar outra questão. Sendo a explicação do comportamento induzido, estudado após a pesquisa.

Tendo como objetivo central investigar a escolha do tipo de oficina por um determinado cliente a partir de uma base de dados de sinistros já ocorridos, foram levantados dados históricos. A extração desta base de dados foi feita pela ferramenta *SQL Server Management Studio* e o tratamento dos dados através do *Microsoft Excel*.

A metodologia utilizada para a abordagem deste problema será o *Cross-Industry Process for Data Mining* (CRISP-DM), idealizada em 1996 pela *European Strategic Program on Research in Information Technology* (ESPRIT) (CHAPMAN, 2000).

Conforme afirma Chapman (2000), é uma robusta abordagem para problemas de análise de dados em empresas. A metodologia pode ser entendida em 6 passos:

1. *Business Understanding* (Entendimento do Negócio)
2. *Data Preparation* (Preparação dos Dados)
3. *Data Understanding* (Entendimento dos Dados)
4. *Modeling* (Modelagem)
5. *Evaluation* (Medição)
6. *Deployment* (Implantação)

Entretanto, o presente trabalho vai se limitar até a quinta etapa, pois o modelo ainda não foi implantado na ferramenta de front-end do portal de abertura de sinistro.

2 Fundamentação Teórica

2.1 Aprendizado de Máquina

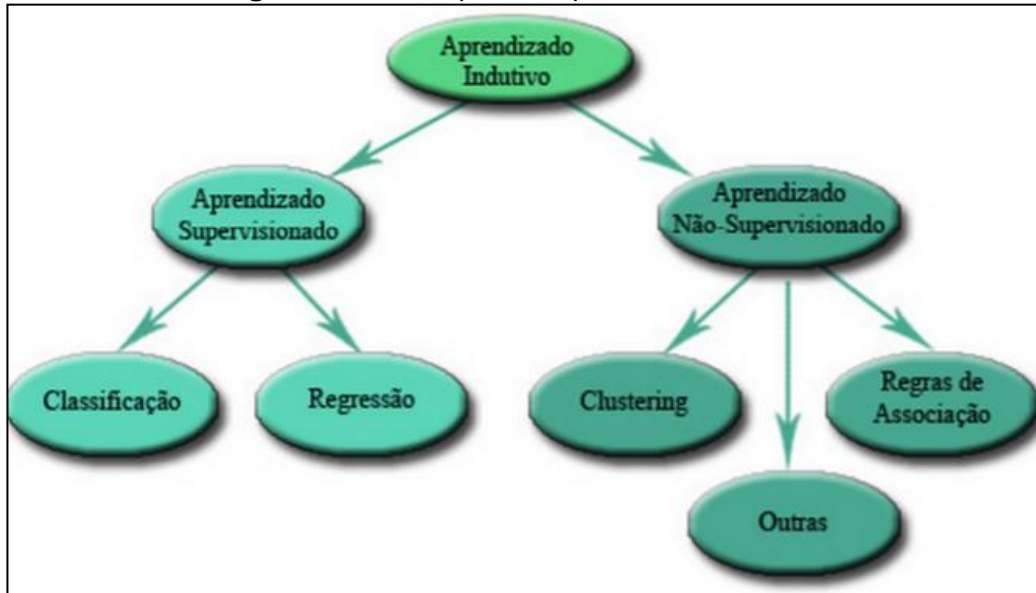
De acordo com Camargo (2007), a Aprendizagem de Máquina (AM) é uma subárea da Inteligência Artificial concentrada em desenvolver modelos que possam “aprender” através da experiência. O aprendizado se dá através de algoritmos dedutivos que baseados em estatística, extraem regras e padrões em grandes massas de dados.

Segundo Rezende (2003), as técnicas de AM empregam um princípio de inferência denominado indução, no qual obtém-se conclusões genéricas a partir de um conjunto particular de exemplos. A indução é caracterizada pelo raciocínio originado em um conceito específico e generalizado, ou seja, da parte para o

todo. Na indução, um conceito é aprendido efetuando-se inferência indutiva sobre os exemplos apresentados. Logo, as hipóteses geradas através da inferência indutiva podem ou não preservar a verdade.

O aprendizado indutivo pode ser dividido em dois tipos principais: Supervisionado e Não-Supervisionado. Conforme a Figura 3 abaixo.

Figura 3 – Hierarquia do Aprendizado Indutivo



Fonte: Rezende, 2003

No aprendizado supervisionado temos a figura de um professor externo, o qual apresenta o conhecimento do ambiente por conjuntos de exemplos na forma: entrada, saída desejada. Logo o algoritmo de AM extrai a representação do conhecimento a partir desses exemplos. Com isso, o objetivo é que a representação gerada seja capaz de produzir saídas corretas para novas entradas não apresentadas previamente.

No aprendizado Não-Supervisionado, não há a presença de um professor, ou seja, não existem exemplos rotulados. Logo, o algoritmo aprende agrupar as entradas submetidas segundo uma medida de qualidade. Essas técnicas são utilizadas principalmente quando o objetivo for encontrar padrões ou tendências que auxiliem no entendimento dos dados.

2.2 Método de Naive Bayes

De acordo com Zhang (2004), o método *de Naive Bayes* consiste em um conjunto de algoritmos de aprendizado supervisionado baseados na aplicação do teorema de Bayes com a suposição "ingênua" de independência condicional entre cada par de recursos, dado o valor da variável de classe. O teorema de

Bayes afirma a seguinte relação, dada a variável de classe y e o vetor de característica dependente x_1 através de x_n , como visto na Equação (1).

$$P(Y|x_1 \dots x_n) = \frac{P(Y)P(x_1 \dots x_n|y)}{P(x_1 \dots x_n)} \quad \text{Eq. (1)}$$

Usando a condição de independência condicional, temos o descrito na Equação (2)

$$P(X_i|Y, X_i, \dots, X_{i-1}, X_{i+1}, \dots, x_n) = P(x_i|y) \quad \text{Eq. (2)}$$

Para todos, esse relacionamento é simplificado resultando na Equação (3)

$$P(Y|x_1 \dots x_n) = \frac{P(y)\prod_{i=1}^n P(x_i|y)}{P(x_1 \dots x_n)} \quad \text{Eq. (3)}$$

Pode-se usar a estimativa de Maximum A Posteriori (MAP) para estimar $P(y)$ e $P(x_i|y)$, sendo o primeiro a frequência relativa das classes no conjunto de treinamento.

2.3 Testes e Validação

A Matriz de Confusão, é uma ferramenta muito usada para avaliações de modelos de classificação em Aprendizado de Máquina. A Matriz de Confusão é uma tabela que mostra as frequências de classificação para cada classe do modelo. Na Matriz tem-se os seguintes conceitos:

- Verdadeiro positivo (*true positive*): ocorre quando no conjunto real, a classe que está se buscando foi prevista corretamente.
- Falso positivo (*false positive*): ocorre quando no conjunto real, a classe que está se buscando prever, foi prevista incorretamente.
- Falso verdadeiro (*true negative*): ocorre quando no conjunto real, a classe que não está se buscando prever, foi prevista corretamente.
- Falso negativo (*false negative*): ocorre quando no conjunto real, a classe que não está se buscando prever, foi prevista incorretamente.

Segue abaixo um exemplo de uma Matriz de Confusão na Figura 4.

Figura 4 – Matriz de Confusão

		Valor Verdadeiro (confirmado por análise)	
		positivos	negativos
Valor Previsto (predito pelo teste)	positivos	VP Verdadeiro Positivo	FP Falso Positivo
	negativos	FN Falso Negativo	VN Verdadeiro Negativo

Fonte: Autores, 2021

Podendo-se realizar algumas análises importantes decorrentes dessa Matriz.

2.3.1 Acurácia

Aqui, tem-se quanto o modelo acertou das previsões possíveis.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} = \frac{\text{previsões corretas}}{\text{todas as previsões}} \quad \text{Eq. (1)}$$

2.3.2 Recall

Tem-se a proporção de positivos foram identificados corretamente. Em outras palavras, quão bom meu modelo é para prever positivos, sendo positivo entendido como a classe que se quer prever.

$$recall = \frac{TP}{TP+FN} \quad \text{Eq. (2)}$$

2.3.3 Precisão

Tem o propósito de identificar se os positivos foram realmente corretos. Em outras palavras, o quão bem o modelo trabalhou.

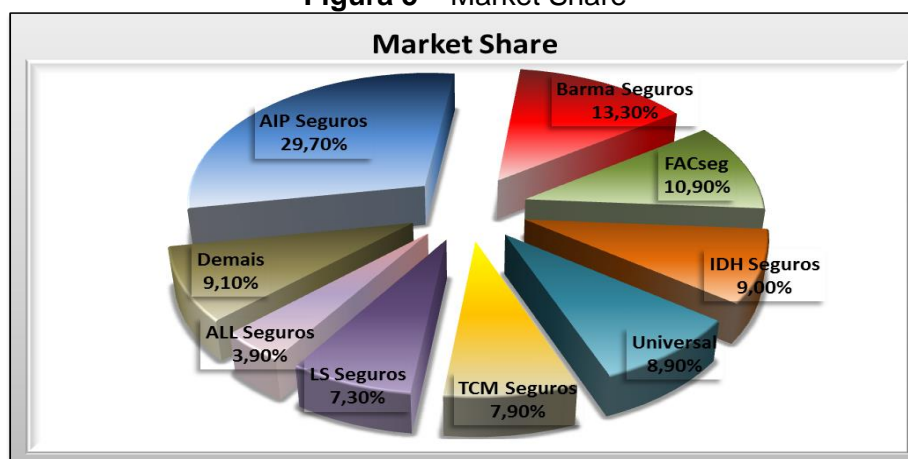
$$precision = \frac{TP}{TP+FP} \quad \text{Eq. (3)}$$

3. Estudo de Caso

O presente estudo foi realizado em uma empresa de seguros de automóveis. A empresa foi fundada em 2004, e especializou-se na operação de seguros de automóveis e ramos elementares, também chamados de seguro patrimonial. Está presente em 70 das 200 maiores empresas do País.

Além de possuir na Carteira de Autos, mais de 1,7 milhão de veículos cobertos, conta com mais de 2 milhões de residências seguradas. A Empresa desenvolve, oferece e administra produtos que são referências no segmento de seguros. Produtos que apresentam flexibilidade na contratação de suas coberturas, proporcionando ótimas alternativas em suas negociações. Com um diferencial importante de agilidade nas cotações, além de grande capacidade de retenção nos riscos que subscreve. Pode-se observar o Market Share da empresa no setor de seguros de automóveis em um nível nacional na Figura 5. Para manter a confidencialidade da empresa em destaque, foram criados nomes fictícios para todas as seguradoras apresentadas na figura abaixo.

Figura 5 – Market Share

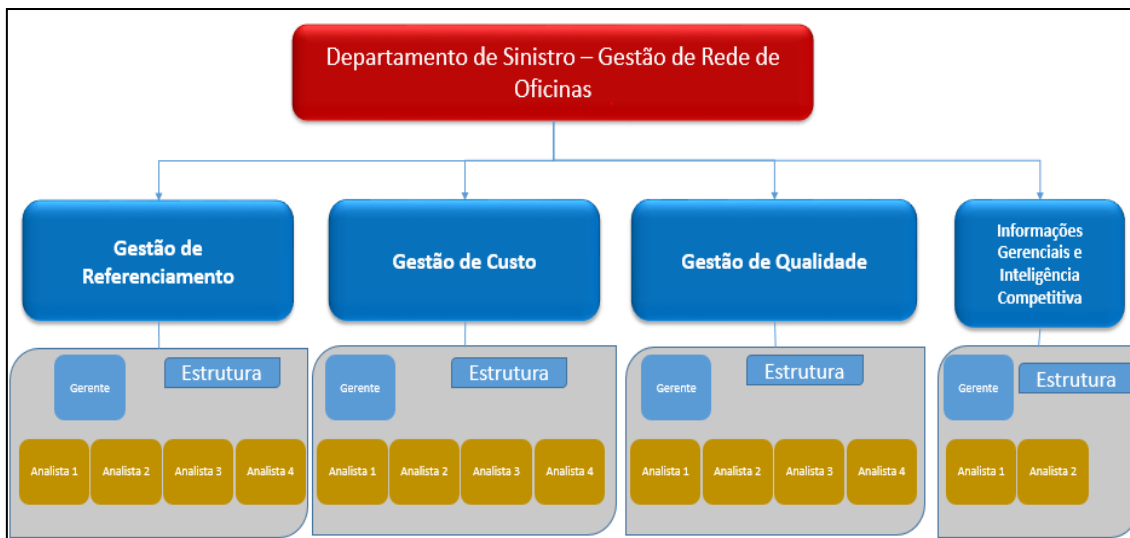


Fonte: Autores, 2021

A FACSEG oferece uma ampla linha de produtos com condições especiais para diversos segmentos, atendendo as necessidades específicas de pessoas físicas e jurídicas, com a relevância de ser uma das maiores seguradoras brasileiras.

O presente estudo será delimitado ao setor de Gestão de Referenciamento de Oficinas, onde tem como atividades principais, basilar, gerenciar e atualizar os cadastros das oficinas que prestam serviços para seguradoras. O setor é subdividido na forma apresentada na Figura 6.

Figura 6 – Organograma da Empresa



Fonte: Autores, 2021

Em uma seguradora de automóveis que tem segurados em todo o território nacional, há um problema que tange o direcionamento dos sinistros para sua rede credenciada de oficinas. Para grande esclarecimento, o segurado ou terceiro tem o livre arbítrio, perante a lei, decidir em qual oficina pretende obter a reparação do seu bem. Existem dois tipos de oficinas: as Oficinas Independentes (particulares) ou Concessionárias, podendo ser Referenciadas ou não Referenciadas.

Para a seguradora é mais interessante o sinistro ser reparado em uma oficina que pertence a sua Rede Referenciada, pois os custos de reparação são mais baixos. As Oficinas Referenciadas possuem mão de obra acordados com essas oficinas. Logo, o direcionamento dos segurados para Rede Referenciada vai impactar diretamente nos custos de reparação, reduzindo o valor total pago, gerando um lucro maior. A Figura 7 demonstra o valor do Custo Médio para cada tipo de oficina.

Figura 7 – Custo Médio por Tipo de Oficina



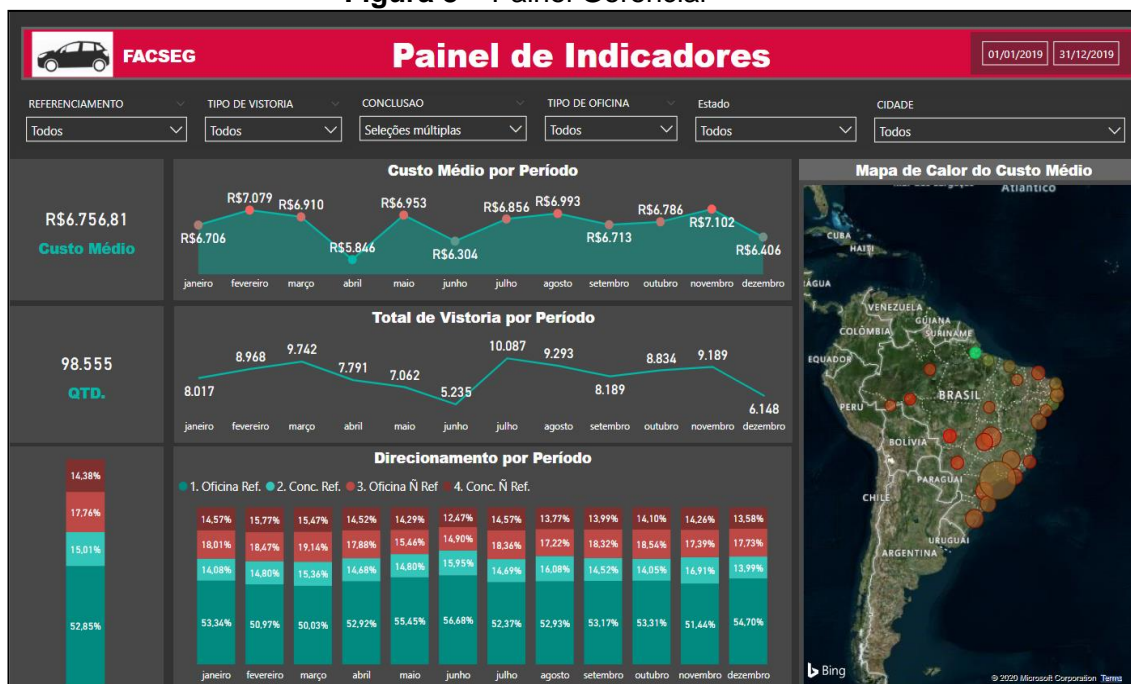
Fonte: Autores, 2021

Como podemos observar, o Custo Médio das Oficinas Concessionárias Não Referenciadas é de R\$ 10.605,18, ou seja, maior. Diante de tal fato, entende-se que, quanto menos automóveis sejam reparados nesta classificação de oficina, é melhor. Por outro lado, o custo da Oficina Independente Referenciada é de R\$ 5.297,10. A diferença de Custo Médio das Concessionárias Não Referenciadas para Independente Referenciada é de R\$

5.484,96. É bom destacar que estamos falando em um cenário nacional, quando escolhemos algumas regiões do país, a discrepância do Custo Médio pode aumentar.

O próximo passo do setor Gestão de Rede da seguradora é entender qual são os principais motivos que levam alguns clientes em preferir Oficinas Não Referenciadas. E assim, apresentar algumas vantagens para os clientes que possam persuadi-los em levar o seu automóvel para uma Oficina Referenciada, gerando maior economia. É importante saber o percentual do direcionamento no decorrer do ano. O direcionamento em uma serie temporal é apresentado na Figura 8.

Figura 8 – Painel Gerencial

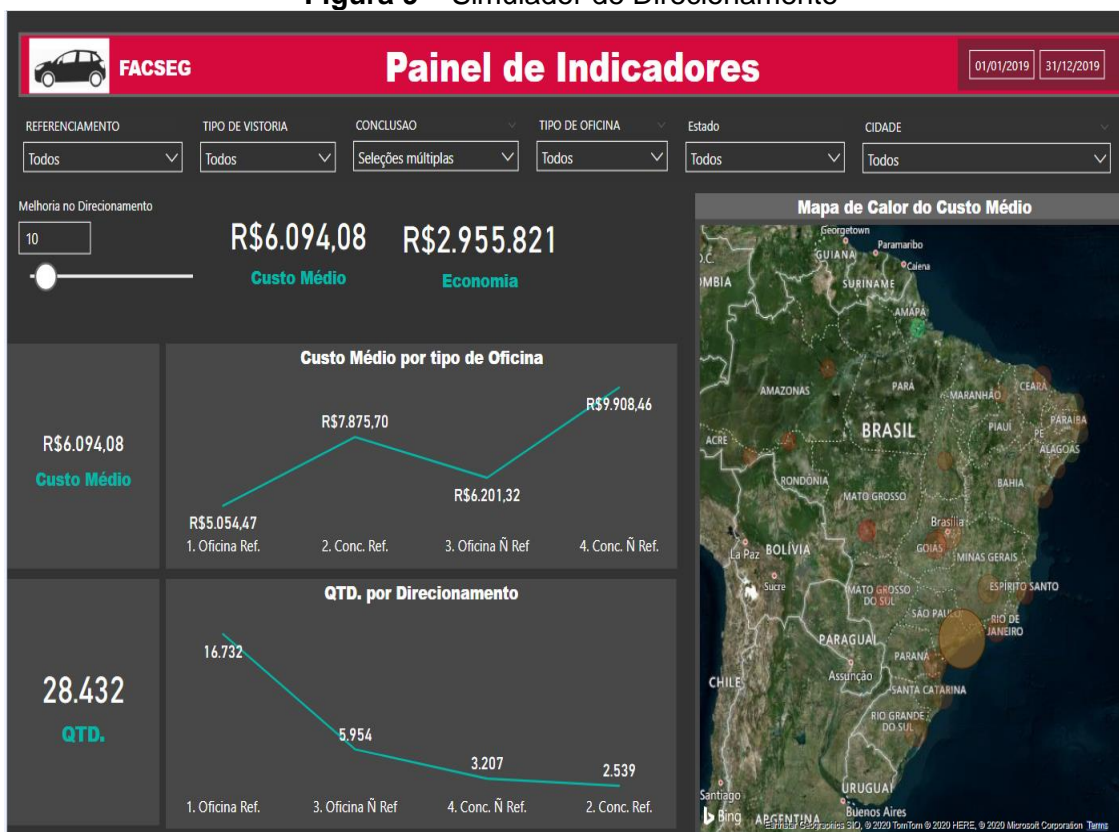


Fonte: Autores, 2021

Considerando a Figura 8, pode-se observar que a média, 52% dos segurados ou terceiros que tem seus veículos sinistrados, preferem levar seus automóveis em uma Oficina Referenciada. Lembrando que, essa categoria de oficina tem o menor Custo Médio, em torno de R\$ 5.200,00. Os outros 48% foram para fora da categoria de menor custo.

O mapa de calor está relacionado com o Custo Médio da Região, enquanto o tamanho do diâmetro da circunferência está relacionado a quantidade de sinistros em um determinado lugar. Sendo assim, pode-se obter como exemplo de análise a região de São Paulo, pois como observado no mapa, tem a maior quantidade de sinistros e não possui Custo Médio inferior. É possível analisar na Figura 9, a diferença de Custo Médio por tipo de oficinas.

Figura 9 – Simulador de Direcionamento



Fonte: Autores, 2021

Observa-se no painel que se os gestores aplicarem um plano de ação para melhorar o direcionamento na região de São Paulo como projeto piloto, baseado no cálculo feito pelos atuários em conjunto com o setor de Gestão de Oficinas Referenciadas, teriam uma economia de quase 3 milhões de reais, caso conseguissem uma performance de 10% no indicador de direcionamento. Vale destacar que, conforme o projeto for amadurecendo, logo os gestores poderão aplicar o algoritmo em todas as regiões do país.

Sendo assim, o presente estudo propõe uso do modelo de *Machine Learning* utilizando o algoritmo de *Naive Bayes*, com o objetivo de conceder um desconto na franquia para os clientes que tem o perfil de não escolher uma Oficina Referenciada. Para tal estudo foi levantado uma amostra com 100 observações diferentes de sinistros abertos, conforme apresenta a Tabela 1.

Tabela 1 - Base de Dados

Status da Idade	Gênero	Status do ano do Automóvel	Origem do Veículo	UF	Classe
Adulto	M	Semi-Novo	IMPORTADO	SP	Sim
Adulto	M	Semi-Novo	NACIONAL	AM	Sim
Adulto	F	Semi-Novo	IMPORTADO	DF	Não
Idoso	M	Novo	NACIONAL	AM	Sim
Adulto	M	Semi-Novo	IMPORTADO	PA	Não
Idoso	M	Semi-Novo	NACIONAL	SP	Não
Idoso	M	Novo	IMPORTADO	SP	Não
Adulto	M	Semi-Novo	IMPORTADO	SP	Não
Adulto	M	Novo	IMPORTADO	SP	Sim
Idoso	M	Semi-Novo	NACIONAL	SP	Não
Jovem	M	Semi-Novo	NACIONAL	SC	Não
Jovem	F	Semi-Novo	NACIONAL	SP	Não
Adulto	M	Usado	NACIONAL	SP	Sim
Adulto	M	Semi-Novo	NACIONAL	SP	Não
Adulto	F	Novo	IMPORTADO	GO	Sim
Idoso	M	Novo	IMPORTADO	MA	Sim
Adulto	F	Semi-Novo	IMPORTADO	SP	Sim

Fonte: Autores, 2021

Os dados acima têm como objetivo aferir a probabilidade de o cliente reparar o seu veículo em uma oficina da Rede Referenciada ou não. A partir dos dados históricos é possível construir um modelo para prever o comportamento de novos clientes. Para o presente estudo estão sendo consideradas as seguintes variáveis.

Origem do Veículo: pode ser Nacional ou Importado.

Nacional são os automóveis que são das seguintes montadoras: GM, VW, FORD e FIAT.

Importados são todos os outros modelos das outras montadoras.

Status da Idade: Pode ser Jovem, Adulto ou Idoso.

Jovem são pessoas de 18 até 30 anos,

Adulto são pessoas de 31 até 50 anos,

Idoso que são pessoas acima de 50 anos.

Gênero: M (Masculino) e F (Feminino).

Status do ano do Automóvel: pode ser Novo, Seminovo ou Usado.

Novos são carros que tem até 3 anos de uso,

Seminovos são carros de 4 anos até 10 anos de uso,

Usados são carros acima de 11 anos de uso.

Classe: se o cliente preferiu levar seu automóvel para uma Rede Referenciada ou não.

No caso, quando estiver marcado como NÃO, o cliente levou o seu automóvel para uma oficina referenciada, não havendo necessidade de conceder desconto na franquia. Quando estiver marcado como SIM, o cliente levou o seu carro em uma oficina que

não é referenciada. Com isso, a proposta é conceder um desconto que leve o cliente para uma oficina referenciada.

A partir dos dados apresentados, pode-se calcular a probabilidade das variáveis de interesses, conforme apresentado na Tabela 2.

Tabela 2 - Variáveis de Interesses

Desconto?	Qtd	%
Sim	48	48%
Não	52	52%
Total	100	100%

Fonte: Autores, 2021

Ou seja, para 48% dos casos será concedido o desconto na franquia e em 52% não há necessidade de conceder o desconto. Além de se ter a probabilidade de ser concedido desconto ou não, é necessário calcular as probabilidades de todos os atributos para cada Variável destacada em cada coluna. Como pode ser observado na Tabela 3.

Tabela 3 - Modelo de dados (Dados)

	Categoria	Sim	Não	Sim	Não
Status Idade	Jovem	5	7	0,1042	0,1346
	Adulto	26	28	0,5417	0,5385
	Idoso	17	17	0,3542	0,3269
Gênero	M	39	31	0,8125	0,5962
	F	9	21	0,1875	0,4038
Status do ano do Automóvel	Novo	25	18	0,5208	0,3462
	Semi-Novos	21	32	0,4375	0,6154
	Usado	2	2	0,0417	0,0385
Origem do Veículo	NACIONAL	24	29	0,5000	0,5577
	IMPORTADO	24	23	0,5000	0,4423
UF	SP	17	18	0,3542	0,3462
	AM	2	0	0,0417	0,0000
	DF	4	3	0,0833	0,0577
	PA	0	1	0,0000	0,0192
	SC	1	2	0,0208	0,0385
	GO	5	1	0,1042	0,0192
	MA	1	1	0,0208	0,0192
	CE	2	1	0,0417	0,0192
	PR	5	2	0,1042	0,0385
	MG	3	3	0,0625	0,0577
	RJ	2	6	0,0417	0,1154
	RS	0	3	0,0000	0,0577
	MT	2	5	0,0417	0,0962
	PE	2	0	0,0417	0,0000
	BA	1	4	0,0208	0,0769
	AL	0	1	0,0000	0,0192
	TO	1	0	0,0208	0,0000
MS	0	1	0,0000	0,0192	

Fonte: Autores, 2021

Após obter o modelo acima, que foi feito por meio dos dados de treino, é necessário analisar os testes de validação. Para isso, foram selecionados 25 clientes para comparar os resultados previstos pelo modelo com o que de fato foi apurado.

Foi utilizado um formulário para analisar o comportamento do algoritmo proposto para o presente estudo. Segue abaixo um exemplo do formulário já com a aplicação do algoritmo.

Figura 10 – Formulário

Formulário		
Idade:	29	Jovem
Sexo:	F	F
Ano do Carro:	2013	Seminovo
Marca do Veículo:	CHEVROLET	NACIONAL
UF	SP	SP
Conceder Desconto na Franquia?	Não	

Fonte: Autores, 2021

Como pode-se observar no exemplo acima, o modelo não concedeu o desconto na franquia, pois classificou que o cliente tem os atributos que aceitaria levar o seu automóvel em uma Oficina Referenciada sem precisar conceder o desconto proposto.

Ainda sobre os testes de validação aplicados sobre os dados de treino pode-se afirmar que Acurácia foi de 88%, ou seja, de cada 10 previsões, o modelo acertou quase 9. Já em relação a Precisão, o modelo apresentou 81,8%, ou seja, de cada 10 clientes que iriam obter desconto, o modelo concedeu assertivamente para 8 pessoas. Em relação ao Recall, o modelo obteve 90% neste índice, ou seja, o modelo tem uma boa capacidade de prever corretamente. É possível visualizar a Matriz da Confusão na Tabela 4.

Tabela 4 - Matriz de Confusão

		Valor Previsto	
		Sim	Não
Valor Verdadeiro	Sim	9	1
	Não	2	13

Fonte: Autores, 2021

4 Considerações Finais

Enfim, retomando o problema de pesquisa: “De qual forma é possível melhorar a performance do indicador de desempenho do direcionamento de sinistro para a rede de Oficinas Referenciada? ” Os dois objetivos da pesquisa foram alcançados: O primeiro objetivo, que consiste em mostrar a aplicação do algoritmo de *Naive Bayes* como método de aprendizagem de máquina no entendimento e modelagem do problema onde foi realizado através das análises dos dados. Foi apresentado um modelo que sugere conceder desconto na franquia ou não para clientes na escolha da oficina para reparação de seu veículo. Já quanto ao segundo objetivo, em avaliar a performance do modelo aplicado, pode-se evidenciar o benefício da utilização de *Machine Learning*, pois foi apresentado bons indicadores de teste e validação como: Acurácia 88%, Precisão 81% e Recall 90%. Sendo assim, pode-se afirmar que foi possível utilizar o *Machine Learning* para modelar o problema da pesquisa.

Enfim, como a quantidade de registros para o treino são pequenos não é possível dizer que os resultados apresentados são conclusivos, mas não é possível dizer que estão incorretos. Assim, os resultados demonstram que o uso de *Machine Learning* são promissores e parecem apontar que esta é uma área promissora a ser explorada.

Vale ressaltar que o modelo ainda não foi implantado para uso do usuário final, pois no presente momento, o setor de Gestão de Oficina está interagindo com o setor de Tecnologia da Informação para colocá-lo no portal de abertura de sinistro.

Para futuras pesquisas, a modelagem por redes neurais poderá ser testada, com uma base enriquecida de informações diretas e indiretas dos clientes, assim como utilizar outros modelos de classificação para analisar o desempenho dos mesmos.

Referências

CAMARGO Y. B. L. de, Abordagem Linguística na Classificação Automática de Textos em Português Tese de Mestrado Programa de Pós-graduação em Engenharia,UFRJ, Mestrado em Ciências em Engenharia Elétrica, Rio de Janeiro, 2007.

CHAPMAN, P. CRISP-DM 1.0: Step-By-Step Data Mining Guide. [S.I.]: 2000.

DATA HACKERS. Entendendo o que é Matriz de Confusão com Python. Disponível em: <<https://medium.com/data-hackers/entendendo-o-que-%C3%A9-matriz-de-confus%C3%A3o-com-python-114e683ec509/>>. Acesso em 15/05/2021.

FREITAS, Ernani Cesar; PRODANOV, Cleber Cristiano. Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico 2. ed. Rio Grande, do Sul, 2013.

MARCONI, M.; LAKATOS, E. M. Metodologia Científica. São Paulo: Editora Atlas, 2004.

OLIVEIRA, M. F. Metodologia científica: um manual para a realização de pesquisas em Administração. Catalão: UFG, 2011.

REZENDE, S.O.; PUGLIESI J.B.; MELANDA, E. A.; PAULA.; Mineração de Dados, in REZENDE,S.O (Eds.), Sistemas Inteligentes, Editora Manole Ltda., p.307-355. 2003.

ZHANG, H. The optimality of naive bayes. In: BARR, V.; MARKOV, Z. (Ed.). Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004). [S.I.]: AAAI Press, 2004.