

Universidade do Minho
Escola de Engenharia

David Garcia Ribeiro

Sugestões para o Retalho de Automóveis Novos: Uma Análise às
Características

Dissertação de Mestrado

Mestrado integrado em Engenharia e Gestão de Sistemas de
Informação

Trabalho realizado sob a orientação do Professor Doutor Jorge
Oliveira e Sá

outubro de 2019

DECLARAÇÃO

Nome: David Garcia Ribeiro

Endereço eletrónico: david.garcia.ribeiro@gmail.com

Telefone: 916265777

Bilhete de Identidade/Cartão do Cidadão: 14077511

Título da dissertação: Sugestões para o Retalho de Automóveis Novos: Uma Análise às Características

Orientador: Professor Doutor Jorge Oliveira e Sá

Ano de conclusão: 2019

Mestrado integrado em Engenharia e Gestão de Sistemas de Informação

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, 17/10/2019

Assinatura

Agradecimentos

Em primeiro lugar quero agradecer à minha família, pais e irmão, e à minha namorada pelo constante apoio, incentivo, disponibilidade, motivação e principalmente pela paciência durante todo este processo.

Ao Professor Doutor Jorge Oliveira e Sá, meu orientador da dissertação, em primeiro lugar por ter aceite a proposta, pela permanente disponibilidade, pelo constante apoio durante todo o processo de desenvolvimento deste projeto e também pela paciência que teve comigo.

Ao Professor Filipe Portela e à aluna de doutoramento Carolina Ribeiro por toda ajuda na resolução das minhas dúvidas principalmente no que diz respeito a questões de *Data mining*.

Este projeto foi um desafio enorme, que me permitiu aprender coisas novas e também alargar o meu conhecimento em outras áreas, exigiu uma enorme quantia de horas de trabalho, mas no fim fez me sentir orgulhoso com o trabalho desenvolvido.

Um grande Obrigado a todos!!

Resumo

Hoje em dia, os consumidores pretendem adquirir veículos com determinadas características e não querem esperar que esses veículos sejam construídos, devido ao *lead-time* da indústria. Assim sendo, as marcas automóveis têm um certo número de veículos em stock com diversas combinações de características tais como cores, jantes e navegação, entre outros, de modo a satisfazer as necessidades dos consumidores, o que nem sempre é possível. Neste projeto apresenta-se uma recolha de dados para perceber quais as características que os consumidores desejam e esses dados foram analisados comparativamente com dados de vendas reais de veículos para validar se as preferências estão de acordo com as vendas. Os dados das vendas reais foram também analisados através de técnicas de mineração com o objetivo de criar modelos que fossem capazes de prever as vendas positivas e negativas, mas principalmente com o objetivo de descobrir algumas regras diferenciadoras em cada Gama, de maneira a ajudar os concessionários a efetuarem encomendas mais acertadas. Após a realização das análises referidas conseguiu-se entender melhor o desejo por certos opcionais em determinadas gamas, nomeadamente o melhoramento das jantes do veículo que é menos valorizado nas Gamas Citadino e SUV/Jipe. Dessa forma, este estudo poderá ajudar as marcas automóveis a adquirir veículos que estejam de acordo com as preferências dos consumidores para que os veículos sejam rapidamente vendidos e os consumidores fiquem satisfeitos.

Palavras chave: Análise de dados; Mineração de dados; Oferta; Procura; Retalho Automóvel;

Abstract

Nowadays, consumers would like to acquire vehicles with certain characteristics and do not want to wait for those vehicles to be built, due to the production lead-times. Therefore, dealerships have a certain number of vehicles in stock with diverse feature combinations such as colours, alloy wheels and car navigation features, among others, with the objective of satisfying the necessities of the consumers, which is not always possible. In this paper, a data collection is presented with the objective of understanding which features are desired by the consumers, that data was comparatively analysed with data from actual sales, to validate if those preferences are aligned with real vehicles sales. Actual sales data was also analysed through *Data mining* techniques with the objective of creating models that are able to predict positive and negative sales, but mainly finding out some differentiating rules in each segment, trying to help dealerships order correctly. After doing said analysis, we were able to better understand the desire features in certain segments, for example, the upgrade of standard wheels it is less valued in the city car and SUV segments. Thus, this study will help dealerships order vehicles to stock, based on consumer preferences with the goal of selling vehicles rapidly and consumers becoming more satisfied.

Key words: Automotive Retail; Data Analysis; *Data mining*; Demand; Supply

Índice

Agradecimentos.....	iii
Resumo.....	v
Abstract.....	vii
Índice.....	ix
Lista de Figuras.....	xii
Lista de Tabelas.....	xiv
1. Introdução.....	1
1.1 Problema.....	1
1.2 Objetivos.....	2
1.3 Estrutura do documento.....	2
2. Metodologia de investigação.....	4
2.1 Metodologia para a realização de entrevistas (Gestores de Vendas).....	4
2.2 Metodologia para a realização do Questionário (Inquiridos).....	4
2.3 Metodologia para a revisão de literatura.....	4
2.4 Metodologia de investigação.....	5
Aplicação da metodologia de investigação.....	7
2.5 Metodologia <i>Data mining</i>	8
3. Revisão de literatura.....	11
3.1 Técnicas de exploração de dados.....	11
3.2 Trabalhos relacionados.....	13
4. Trabalho realizado.....	15
4.1 Dados dos consumidores.....	16
4.2 Dados dos concessionários.....	19
4.3 <i>Data mining</i>	21
Classificação.....	24
5. Discussão de resultados.....	25
5.1 Resultados da análise manual.....	25
5.1.1 Comparação por gamas da quantidade de veículos preferidos pelos consumidores com os vendidos pelos concessionários.....	25
5.1.2 Análise aos opcionais que se podem comparar com os dados do concessionário.....	25
5.1.3 Análise aos dados dos concessionários através das combinações obtidas no questionário.....	26

5.1.4	Análise aos tempos médios das combinações mais vendidas pelos concessionários.....	26
5.2	Análise com técnicas de <i>Data mining</i>	28
5.2.1	Equilíbrio das classes dos Targets.....	28
5.2.2	Resultados da Classificação	34
5.2.3	Modelo que obteve melhores resultados para cada Gama	53
5.2.4	Sugestões para cada Gama	55
6.	Conclusões e Limitações.....	57
6.1	Conclusões e Limitações	57
6.2	Trabalho Futuro.....	58
	Bibliografia	59
	Anexo I	62
	Anexo II	69
	Anexo III	74
	Anexo IV	81
	Anexo V	86

Lista de Figuras

Figura 1: Metodologia DSR.....	5
Figura 2: Metodologia CRISP-DM.....	9
Figura 3: Equilíbrio das classes nos diferentes Targets para a gama Cidadino.....	29
Figura 4: Equilíbrio das classes nos diferentes Targets para a gama Desportivo.....	30
Figura 5: Equilíbrio das classes nos diferentes Targets para a gama Familiar.....	31
Figura 6: Equilíbrio das classes nos diferentes Targets para a gama Luxo/Supercarro.....	32
Figura 7: Equilíbrio das classes nos diferentes Targets para a gama SUV/Jipe.....	33

Lista de Tabelas

Tabela 1 – Género	16
Tabela 2 – Idades	16
Tabela 3 - Grau de escolaridade	17
Tabela 4 – Profissão	17
Tabela 5 - Classificação da situação financeira	17
Tabela 6 - Listagem dos opcionais sobre que os consumidores foram inquiridos	18
Tabela 7 – Valores dos Opcionais (1), (2) e (4)	18
Tabela 8 – Combinações de características - dados do questionário	19
Tabela 9 - Combinações de vendas - dados do concessionário	20
Tabela 10 - Combinações de variáveis nos diferentes cenários	23
Tabela 11 - Diferentes valores ocupados pelas classes nas diferentes	23
Tabela 12 – Cores analisadas e a sua representação nos dados	23
Tabela 13 – Preferências versus vendas por Gama	25
Tabela 14 - Dados do consumidor com os opcionais comparáveis	26
Tabela 15 - Combinações de características de vendas do concessionário	26
Tabela 16 - Análise de vendas por gama	27
Tabela 17 – Resultados da análise aos dados do concessionário na gama citadino para o Target T1	34
Tabela 18 – Resultados da análise aos dados do concessionário na gama Desportivo/Cabrio para o Target T1	37
Tabela 19 – Resultados da análise aos dados do concessionário na gama Desportivo/Cabrio para o Target T3	39
Tabela 20 – Resultados da análise aos dados do concessionário na gama Familiar para o Target T1	42
Tabela 21 – Resultados da análise aos dados do concessionário na gama Luxo/Supercarro para o Target T1	45
Tabela 22 – Resultados da análise aos dados do concessionário na gama Luxo/Supercarro para o Target T4	47
Tabela 23 – Resultados da análise aos dados do concessionário na gama SUV/Jipe para o Target T1	50
Tabela 24 – Resultados da análise aos dados do concessionário na gama citadino para o Target T2	62
Tabela 25 – Resultados da análise aos dados do concessionário na gama citadino para o Target T3	64
Tabela 26 – Resultados da análise aos dados do concessionário na gama citadino para o Target T4	66
Tabela 27 – Resultados da análise aos dados do concessionário na gama Desportivo/Cabrio para o Target T2	69
Tabela 28 – Resultados da análise aos dados do concessionário na gama Desportivo/Cabrio para o Target T4	71
Tabela 29 – Resultados da análise aos dados do concessionário na gama Familiar para o Target T2	74
Tabela 30 – Resultados da análise aos dados do concessionário na gama Familiar para o Target T3	76
Tabela 31 – Resultados da análise aos dados do concessionário na gama Familiar para o Target T4	78
Tabela 32 – Resultados da análise aos dados do concessionário na gama Luxo/Supercarro para o Target T2	81
Tabela 33 – Resultados da análise aos dados do concessionário na gama Luxo/Supercarro para o Target T3	83
Tabela 34 – Resultados da análise aos dados do concessionário na gama SUV/Jipe para o Target T2	86
Tabela 35 – Resultados da análise aos dados do concessionário na gama SUV/Jipe para o Target T3	88
Tabela 36 – Resultados da análise aos dados do concessionário na gama SUV/Jipe para o Target T4	90

1. Introdução

1.1 Problema

Os concessionários debatem-se com o problema de, periodicamente, serem obrigados pelo produtor ou importador a fazerem encomendas de veículos para stock. Apesar de, baseado em experiência de vendas passadas, conseguirem acertar nas quantidades certas a encomendar por Gamas, Cores e opcionais, fazendo com que os veículos fiquem pouco tempo em stock, por vezes certos veículos com determinadas combinações de características ficam demasiado tempo em stock, problema que foi identificado em discussões informais com gestores de vendas da área. Por outro lado, os consumidores desejam adquirir veículos com determinadas características; quando tal acontece, o consumidor dirige-se a um ou múltiplos concessionários na procura do seu automóvel e nem sempre encontra o veículo que pretende.

Analisando o lado do consumidor, frequentemente, o consumidor não encontra o veículo que deseja disponível de imediato, como opção pode esperar pelo veículo ou escolhe um veículo que esteja disponível em stock. Para tentar perceber melhor as preferências dos consumidores, este trabalho apresenta um levantamento, baseado num questionário, das combinações de características preferidas pelos consumidores, tais como: gamas, cores exteriores, interiores e ainda 26 opcionais.

Analisando agora o lado do concessionário, verifica-se que há veículos em stock que demoram muitos dias a serem vendidos, o que pode acontecer devido à combinação das características do veículo não ser do agrado dos consumidores. Para tentar perceber esse problema, recolheram-se dados de uma rede de concessionários portugueses representantes de várias marcas automóveis com 5.021 registos de vendas de veículos.

Posteriormente foi realizada uma análise comparativa (manual) entre as duas fontes de dados para validar se as preferências dos consumidores estão de acordo com as vendas dos concessionários, facilitando depois o exercício seguinte.

Depois de se ganhar algum conhecimento sobre os dados e de se perceber algumas diferenças entre os subgrupos selecionados (gamas), procedeu-se a uma análise dos dados da rede de concessionários, através de técnicas de mineração de dados, com recurso a algoritmos de classificação, com vista a procurar modelos que conseguissem classificar as vendas de cada subgrupo em positivas ou negativas, isto é, as vendas que forem efetuadas no intervalo de dias mais perto de zero, são consideradas como vendas positivas e as que forem efetuadas no intervalo

de dias mais distante de zero ou mais próximo de infinito, são consideradas vendas negativas, procurando assim modelos que conseguissem diferenciar e prever os dois tipos de vendas, mas principalmente com o objetivo de procurar uma ou várias regras para cada Gama que ajudem os concessionários a encomendar os veículos certos, isto é, que sejam facilmente vendáveis.

1.2 Objetivos

Procura-se ajudar os concessionários a satisfazerem melhor a procura ao encomendarem veículos que satisfazem os desejos dos consumidores. Primeiramente, tentaram-se obter as melhores combinações de cores e equipamentos dentro de cada Gama, através de uma análise manual comparativa dos dois conjuntos de dados. Em segundo lugar procuram-se modelos que consigam prever vendas positivas ou negativas, utilizando técnicas de *Data mining* com algoritmos de classificação.

Mas o intuito principal é procurar alguma regra, ou conjunto de regras que seja capaz de alguma forma aumentar a satisfação dos envolvidos, no sentido de tornar mais fácil as encomendas, pelo menos dentro de cada gama, aumentando assim os níveis de satisfação, tanto na produção e venda do produto como do consumidor.

1.3 Estrutura do documento

Este documento foi estruturado em seis capítulos, a saber:

Capítulo 1: Este é o capítulo introdutório, onde é apresentado o problema e os objetivos referentes a este projeto.

Capítulo 2: Neste capítulo são abordadas as diferentes metodologias de investigação, para as entrevistas, a realização do questionário, a revisão de literatura, a investigação e para o *Data mining*.

Capítulo 3: Neste capítulo é feita a revisão fundamental e ainda o estado da arte; para além de serem recolhidos os fundamentos também se recolheu os avanços já conhecidos nas áreas que foram abordadas.

Capítulo 4: Neste capítulo são apresentadas as duas diferentes recolhas de dados, o questionário aos consumidores e os dados de vendas reais recolhidos no concessionário, assim como as transformações que sofreram de forma a assistir as análises no capítulo seguinte.

Capítulo 5: Neste capítulo são apresentados os resultados das duas análises efetuadas aos dados, a manual onde foram comparados os desejos dos consumidores com as vendas efetivas do concessionário e ainda, através de técnicas de mineração, foram procurados modelos que fossem capazes de classificar as vendas já efetuadas em positivas ou negativas segundo o tempo que demoraram a vender, para apoiar esta última tarefa foram criados 15 cenários e 4 Targets diferentes

Capítulo 6: Neste capítulo são apresentadas as conclusões, todas as limitações que foram identificadas durante a realização do trabalho e ainda o trabalho futuro a realizar com objetivo de obter resultados mais fidedignos.

2. Metodologia de investigação

Neste capítulo são apresentadas as diferentes metodologias abordadas para a realização deste projeto. Primeiro começa-se pela metodologia para a realização de entrevistas aos Gestores de Vendas, a seguir apresenta-se a metodologia para a realização do questionário (Inquiridos) e ainda a metodologia para a realização da revisão de Literatura. Como metodologia de investigação utilizamos o Design Science Research e mostra-se a sua aplicação e em último lugar temos a metodologia utilizada para o *Data mining* (CRISP-DM).

2.1 Metodologia para a realização de entrevistas (Gestores de Vendas)

Este processo remonta à altura em que foi proposto este trabalho de dissertação. Foram feitas 2 entrevistas, o mais similares possível, uma escrita e outra oral, completamente informais, pois na altura apenas se procurava obter mais conhecimento sobre a área e sobre os processos que ocorrem, que normalmente não estão à vista do público geral. Os resultados dessas entrevistas estão presentes no início do capítulo 3.

2.2 Metodologia para a realização do Questionário (Inquiridos)

Para a realização de questionários procurou-se perceber o que já tinha sido feito por outros investigadores, recolha que está explicada no segundo capítulo deste documento. Também se teve o cuidado de testar o questionário numa população de teste (10 pessoas) antes do seu Deployment, com o objetivo de evitar dúvidas a quem vai responder ao questionário.

2.3 Metodologia para a revisão de literatura

A revisão de literatura foi feita com objetivo de perceber melhor a área de investigação, perceber o que já foi feito e entender os conceitos fundamentais, assim como a recolha de metodologias utilizadas por outros em investigações similares. O objetivo principal é ter um conhecimento o mais alargado possível sobre a área de forma a tentar acrescentar algo à área que se está a investigar. Para atingir os objetivos propostos, foi realizada uma recolha de resultados de investigações já realizadas na área e também foram recolhidos os conceitos fundamentais para esta investigação, visando assim promover uma investigação completa e adequada.

Para a recolha de resultados de outras investigações e também dos conceitos fundamentais foram utilizados motores de busca online especializados na procura de publicações científicas, assim como o google scholar, scopus e web of science. Algumas publicações científicas foram ainda encontradas na bibliografia de outras publicações científicas.

A identificação e a seleção das publicações científicas utilizadas foram realizadas tendo por base as seguintes palavras chave: *vehicle preferences survey, consumer preferences buying cars, vehicle feature preference, Data mining automotive industry, Data mining automotive retail* e ainda para as abordagens metodológicas: *design science research in information systems* e CRISP-DM, entre outras que não apresentaram resultados tão significativos. Tentou-se sempre utilizar informação o mais recente possível, apesar de nem sempre ser possível.

2.4 Metodologia de investigação

A metodologia de investigação empregue neste projeto, será o Design Science Research (DSR), tendo em conta os objetivos e a finalidade definidos anteriormente.

O DSR é uma alternativa ou um complemento do comportamento natural que é a abordagem científica, que é dominante na pesquisa de Sistemas de Informação. Com DSR, o investigador “cria e avalia artefactos de sistemas de informação, com o objetivo de resolver determinados problemas organizacionais”. Envolve a criação de artefactos inovadores de tecnologias de informação - artefactos usados para resolver problemas que afetam a prática de TI (Arnott & Pervan, 2012).

Tendo em conta os objetivos e finalidade deste projeto, segue-se o modelo apresentado na figura 1:

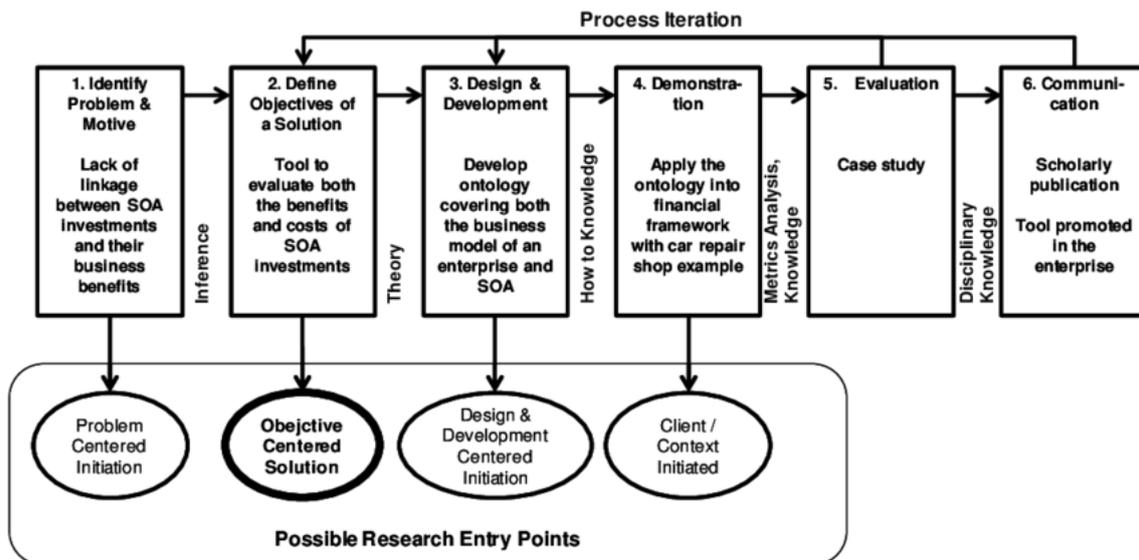


Figura 1: Metodologia DSR (imagem retirada do documento “Design science in decision support systems research: An assessment using the Hevner, March, Park, and Ram Guidelines” (Arnott & Pervan, 2012))

Identificação do problema e motivação

Definir um problema de pesquisa específico e justificar o valor de uma solução. Visto que a definição do problema será usada para desenvolver o artefacto que efetivamente consegue entregar uma solução, pode ser que seja útil atomizar o problema conceptualmente para que a solução possa capturar a sua complexidade. Justificar o valor de uma solução resulta em duas situações, motiva o investigador e a audiência da pesquisa em busca de uma solução e para aceitar os resultados ajuda a perceber as razões associadas com o entendimento do problema por parte do investigador (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007).

Definir os objetivos da solução

Inferir os objetivos de uma solução da definição do problema e do conhecimento do que é possível. Os objetivos podem ser quantitativos, como termos em que a solução desejável será melhor do que alguma atual, ou qualitativo, como uma descrição de como este novo artefacto suporta soluções a problemas não resolvidos até agora. Os objetivos têm de ser inferidos racionalmente da especificação do problema (Peffer et al., 2007).

Design e Desenvolvimento

Criar o artefacto. Estes artefactos são potencialmente, construtores, modelos, métodos ou instanciações. Conceptualmente, um artefacto de Design Research pode ser qualquer objeto designado em que uma contribuição está embebida no design. Esta atividade inclui determinar a funcionalidade e arquitetura desejada de um artefacto e depois criar o artefacto. Os recursos necessários para avançar dos objetivos para o design e desenvolvimento incluem conhecimento de teoria que pode ser trazida numa nova solução (Peffer et al., 2007).

Demonstração

Demonstrar o uso do artefacto para resolver uma ou mais instâncias do problema. Isto pode envolver o seu uso em experimentação, simulação, caso de estudo, prova ou outra qualquer atividade apropriada. Os requisitos necessários para a demonstração incluem conhecimento efetivo de como se utiliza o artefacto para resolver o problema (Peffer et al., 2007).

Avaliação

Observar e medir quão bem o artefacto suporta a solução do problema. Esta atividade envolve comparar os objetivos de uma solução a resultados observados atualmente do uso deste artefacto

na demonstração. Requer conhecimento de métricas relevantes e técnicas de análise. Dependendo da natureza do foro do problema e do artefacto, a avaliação pode assumir muitas formas (Peffer et al., 2007).

Comunicação

Comunicar o problema e a sua importância, o artefacto, a sua utilidade, o rigor do seu design e a sua eficácia para investigadores e outras audiências relevantes. Em publicações escolares, os investigadores podem utilizar a estrutura do processo para estruturar o artigo. A comunicação requer conhecimento da cultura disciplinar (Peffer et al., 2007).

Aplicação da metodologia de investigação

Identificação do problema e motivação

Nesta fase definimos o problema, presente na primeira parte da introdução deste documento, em resumo, pode-se concluir que os concessionários não conseguem satisfazer os consumidores com os veículos que dispõem em stock. Também nesta fase é importante perceber o trabalho feito anteriormente, assim como a revisão fundamental e o estado da arte presentes no segundo capítulo deste documento. A definição do problema e entendimento do trabalho previamente realizado permite a obtenção de conhecimento para a realização das próximas etapas.

Definir os objetivos da solução

Nesta fase definem-se os objetivos da solução, já presentes na segunda parte da introdução a este documento, na secção “Objetivos”. Em que se pretende obter as melhores combinações dentro de cada Gama através da análise manual, classificar as vendas do concessionário em positivas ou negativas tendo em conta o tempo que demoraram a vender e principalmente determinar um conjunto de regras que sirvam de apoio à encomenda de veículos novos por parte dos gestores de vendas dos concessionários.

Design e Desenvolvimento

O Design e desenvolvimento de uma solução está presente no capítulo 3 deste documento e está dividido em 3 fases. A primeira fase passou pela construção, desenvolvimento e difusão de um questionário de forma a perceber melhor os desejos dos consumidores. A segunda fase passou

pela recolha e análise manual a dados recolhidos em concessionários. A terceira fase, utilizou os dados recolhidos na segunda fase para, através de técnicas de mineração de dados, procurar modelos que conseguissem prever melhor as vendas e principalmente para perceber a influência das diferentes variáveis na solução final.

Demonstração

Esta fase está presente no capítulo 4 deste documento, na discussão de resultados, onde se fazem várias avaliações manuais a ambos os conjuntos de dados recolhidos e onde ainda se aplicam diferentes algoritmos aos dados recolhidos no concessionário.

Avaliação

Como são utilizadas múltiplas fontes (dados do questionário e do concessionário), a validação é uma combinação dessa mistura de análises, que permite de certa forma confirmar que os resultados estão corretos através da comparação com os dados das recolhas, podemos então afirmar que a própria comparação é uma validação dos dados. O que foi inicialmente proposto foi a realização de um Focus Group, que é uma discussão informal entre um certo grupo selecionado de indivíduos, sobre um tópico em específico. Uma das forças principais da metodologia Focus Group é a flexibilidade e as diferentes maneiras de uso. Quando conduzido com um Framework epistemológico essencialista implica a assunção que os indivíduos têm as suas ideias pessoais, opiniões e entendimentos (Wilkinson, 1998). Considerou-se que um Focus Group com alguns gestores de vendas, funcionaria bem como forma de validação dos resultados obtidos, mas devido a restrições de tempo, por se tratar de um trabalho de mestrado, não foi possível chegar a esta fase.

Comunicação

Nesta última etapa, o principal objetivo é permitir a gestores de vendas da área encomendar veículos que satisfaçam melhor os consumidores, mas também dar a conhecer a outros investigadores o trabalho que foi feito e as soluções que derivaram dele, assim como todas as limitações que surgiram durante a realização do projeto.

2.5 Metodologia *Data mining*

O Cross-Industry Standart Process for *Data mining* (CRISP-DM) é uma metodologia popular para aumentar o sucesso de projetos de *Data mining* (Chapman et al., 2000). A metodologia define uma sequência de seis fases não rígidas, que permitem a construção e implementação de um

modelo de DM para ser usado num ambiente real, ajudando a suportar decisões de negócio (Moro, Laureano, & Cortez, 2011), metodologia que está dividida em seis fases, apresentadas na figura 2:

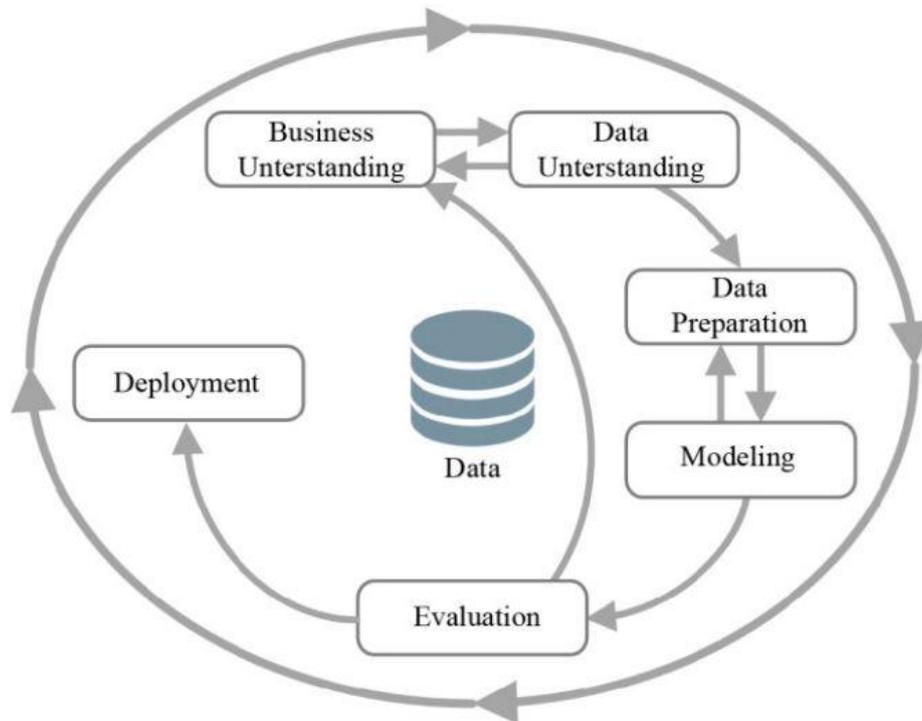


Figura 2: Metodologia CRISP-DM (imagem retirada do documento: “DMME: *Data mining* methodology for engineering application – an holistic extension to the CRISP-DM model” (Huber, Wiemer, Schneider, & Ihlenfeldt, 2019))

O CRISP-DM define um projeto como um processo cíclico, onde diversas iterações podem ser utilizadas para permitir um resultado final mais afinado no sentido dos objetivos do negócio. Depois de identificar a meta a atingir (Business Understanding Phase), os dados precisam de ser analisados (Data Understanding) e processados (Data Preparation). A fase de Modelação (Modeling) constrói o modelo que representa o conhecimento adquirido (e.g. dada uma instância, o modelo pode ser usado para prever o valor do target que presente a meta definida). De seguida, o modelo é analisado na fase de Avaliação (Evaluation), em termos de desempenho e utilidade. Por exemplo, em tarefas de classificação (para prever um target discreto), as métricas comuns são as da Matriz de Confusão (Foster Provost & Kohavi, 1998) e a curva Receiver Operating Characteristic (ROC) (Fawcett, 2006). Se o modelo obtido não é bom que chegue para ser utilizado como suporte ao negócio, então é definida uma nova iteração para o CRISP-DM. Senão, o modelo é implementado no ambiente de tempo real (Deployment Phase) (Moro et al., 2011).

No projeto desenvolvido aplicou-se o CRISP-DM com o objetivo de orientar e aumentar o sucesso deste projeto de *Data mining*. Na primeira fase, de compreensão do negócio (Business Understanding), percebemos os objetivos e as motivações, já presentes no primeiro capítulo deste documento e foi feita ainda uma recolha de como funcionam estes processos no estrangeiro, presente no início do capítulo 4. Na segunda fase de compreensão de dados, (Data Understanding), foi feita uma descrição dos dados e uma análise manual aos mesmos, que está presente no quarto capítulo deste documento. Na terceira fase, preparação de dados (Data Preparation), foram feitas transformações aos dados que estão explicadas durante o capítulo 4 deste documento. Na quarta fase, de Modelação (Modeling), são criados Targets e cenários e são escolhidos os algoritmos que se vão aplicar aos dados, presente na última parte do capítulo 4. Na penúltima fase, de Avaliação (Evaluation), são avaliados os resultados obtidos para os 3 diferentes algoritmos aplicados, segundo os diferentes, targets e cenários, presentes na segunda secção do capítulo 5. A última fase, a Implementação (Deployment), não se aplicou neste projeto, devido a ser do âmbito académico, as restrições de tempo não o permitiram, mas ficará de certeza guardado como projeto de trabalho futuro.

3. Revisão de literatura

3.1 Técnicas de exploração de dados

Data mining, geralmente definido como o processo de descoberta de padrões significativos em grandes quantidades de dados, oferece uma boa variedade de técnicas, métodos e ferramentas para uma análise minuciosa dos dados disponíveis em vários campos (Kabakchieva, 2012). Os tipos principais de tarefas de *Data mining* podem ser categorizados como classificação, regressão, *clustering* e *association rule mining*. A tarefa de classificação é possivelmente a tarefa de *Data mining* mais estudada (Wan, 2018). Consiste na construção de um modelo de classificação ou classificador para prever o rótulo da classe (um valor nominal ou categórico) de uma instância através do uso de valores de outras características (atributos de predição) dessa instância (Freitas, 2003; Jiawei Han & Pei, 2011). Na realidade, a essência do processo de classificação é explorar as correlações entre as características e os rótulos de classe das instâncias de forma a encontrar a fronteira entre os rótulos no espaço de dados – um espaço onde a posição de uma instância é determinada pelos valores das características nessa instância (Wan, 2018).

No texto que se segue, apresentam-se três algoritmos que são de uso comum em tarefas de classificação:

J48 – o classificador J48 é uma simples árvore de decisão C4.5 para classificação. Cria uma árvore binária. A abordagem da árvore de decisão é mais útil em problemas de classificação. Com esta técnica a árvore é construída para modelar o processo de classificação (Patil & Sherekar, 2013). Uma vez que a árvore está construída, é aplicada a cada tuplo na base de dados e resulta em classificação para esse tuplo (Dunham, 2006; Sharma & Sahni, 2011).

Naives Bayes – o algoritmo Naives Bayes é um simples classificador probabilístico que calcula um conjunto de probabilidades através da contagem da frequência e das combinações de valores num dado *dataset*. O algoritmo utiliza o teorema de Bayes e assume que todos os atributos são independentes dado o valor da variável da classe (Patil & Sherekar, 2013). Essa condição de suposição de independência raramente é verdade em aplicações no mundo real, daí a caracterização como Naive ainda que o algoritmo tende a desempenhar-se bem e a aprender rápido em vários problemas de classificação (Dimitoglou, Adams, & Jim, 2012).

Logistic Regression – o termo regressão é definido como uma análise ou medição da relação entre uma variável dependente e uma ou mais variáveis mais independentes. A regressão pode

ser definida em dois tipos: regressão linear e regressão logística. A regressão logística é basicamente utilizada para classificar dados de baixa dimensão com limites não lineares. Também fornece a diferença na percentagem de variáveis dependentes e fornece uma classificação das variáveis individuais de acordo com a sua importância. Então, o principal motivo para a regressão logística é determinar o resultado de cada variável corretamente (Sahoo & Kumar, 2012).

Em estatística ou em *Data mining*, uma das tarefas principais é treinar um modelo através dos dados disponíveis. Esses modelos podem ser de regressão ou classificação. Ainda que esses modelos possam mostrar capacidades de predição adequadas nos dados de treino, ao mesmo tempo, podem não prever os dados futuros corretamente. Isto é um problema que precisa de ser resolvido. Cross-Validation (CV) (Allen, 1974; Geisser, 1975; Stone, 1974) é um procedimento para estimar o desempenho geral neste contexto. Cross-Validation é o método mais comumente utilizado para avaliação do desempenho preditivo de um modelo, dado em antemão ou quando é desenvolvido por um modelo de predição. Os dados são normalmente divididos em duas partes e baseado nesta divisão, numa parte é feito o treino enquanto o desempenho de previsão é testado na outra parte (Yadav & Shukla, 2016). Portanto, Cross-Validation é amplamente aceite nas comunidades de *Data mining* e machine learning, e serve como procedimento padrão para o bem da seleção do modelo ou da seleção dos processos de modelação (Hastie, Tibshirani, Friedman, & Franklin, 2005; Zhang & Yang, 2015). O fator mais crítico na aplicação de Cross-Validation é a escolha do rácio da divisão para os dados de treino e de teste (Yadav & Shukla, 2016). Este rácio também é conhecido como tamanho de validação.

Depois da escolha dos algoritmos e do método de divisão dos dados, é necessário escolher as métricas pelas quais os diferentes modelos vão ser avaliados. Uma forma simples de medir o desempenho de um modelo na tarefa de classificação é calcular o rácio de acertos em relação ao número total de registos, podendo-se designar esta medida de acuidade (Marques, Correia, & Cortez, 2008).. Apesar de muitas destas métricas poderem ser calculadas através da matriz de Confusão, usualmente o software de *Data mining* efetua o cálculo dessas métricas automaticamente sem ser necessário intervenção humana. A precisão é usada para medir os padrões positivos que são corretamente previstos do total de padrões previstos na classe positiva (Hossin & Sulaiman, 2015). A especificidade é a métrica que é usada para medir a fração de padrões negativos que são classificados corretamente (Hossin & Sulaiman, 2015) e a sensibilidade é a métrica que é usada para medir a fração de padrões positivos que são classificados corretamente (Hossin & Sulaiman, 2015). Um gráfico de receiver operating characteristics (ROC)

é uma técnica para visualização, organização e seleção de classificadores baseado no seu desempenho (Fawcett, 2006). Nos últimos anos tem-se visto um aumento no uso de gráficos ROC na comunidade de machine learning, devido à percepção de que a simples acuidade da classificação é frequentemente uma métrica pobre para medir o desempenho (F. Provost, Fawcett, & Kohavi, 1998; F. J. Provost & Fawcett, 1997).

Qualquer *dataset* com uma distribuição de classes desigual é tecnicamente desequilibrado (*imbalanced*). Contudo, um *dataset* é dito desequilibrado quando há uma significante, ou em alguns casos extremos, desproporção entre o número de exemplos da classe do problema. Noutras palavras, o desequilíbrio de classes acontece quando o número de exemplos que representa uma classe é muito mais baixo que nas outras classes. Consequentemente, uma ou mais classes podem estar fracamente representadas no *dataset*. Como consequência, as instâncias da classe em minoria são mais frequentemente mal classificadas do que aquelas que estão na de maioria (He & Garcia, 2008).

3.2 Trabalhos relacionados

O sucesso e o fracasso das cadeias de fornecedores são determinados no mercado pelo consumidor final. Obter o produto certo, pelo preço certo, no tempo certo não é apenas o eixo central para o sucesso competitivo, mas também a chave para a sobrevivência (Mason-Jones, Naylor, & Towill, 2000).

Há muitos fatores que condicionam a escolha dos consumidores na compra de um veículo. De maneira a aumentar as vendas, os produtores têm de conhecer as preferências dos consumidores. Por exemplo, a cor do veículo é um fator muito importante porque está diretamente relacionado com o que o consumidor vê (Satake et al., 2011)(Sliburyte & Skeryte, 2014).

A recolha de dados de preferências de consumidores de veículos foi abordada em diversos estudos. No estudo proposto por (Lave & Train, 1979) foram analisadas as preferências de 541 compradores de carros novos, dentro de 10 categorias diferentes divididas por gama, onde se verificou que pessoas mais velhas e agregados familiares com maiores rendimentos têm tendência a escolher veículos de gama superior e mais caros. No estudo proposto por (Kitamura, Golob, Yamamoto, & Wu, 1999) realizou-se um estudo a 1.898 agregados familiares, com seis gamas de veículos em análise, em que se conseguiu apurar que os veículos familiares são escolhidos em áreas com elevado trânsito e os veículos desportivos em áreas com mais densidade residencial. O estudo proposto por Choo & Mokhtarian (Choo & Mokhtarian, 2004) foram inquiridos 1.904

participantes num inquérito com nove alternativas baseadas na gama do veículo e descobriram que a atitude, a personalidade e o estilo de vida são muito importantes na escolha da gama do veículo. Baltas & Saridakis (Baltas & Saridakis, 2013) deram grande ênfase à divisão dos veículos em gamas.

O *Data mining* é um processo iterativo em que o progresso é definido pela descoberta, através de métodos automáticos ou manuais. O *Data mining* é mais útil num cenário de análise exploratória onde não há noções predeterminadas sobre o que é que irá constituir um resultado “interessante”. O *Data mining* é a procura de informação nova, valiosa e não trivial em grandes volumes de dados. É um esforço cooperativo entre homens e computadores. Os melhores resultados são alcançados através do equilíbrio do conhecimento entre humanos especialistas na descrição de problemas e as capacidades de busca dos computadores. A raiz do problema é o tamanho e as dimensões dos dados que são demasiado grandes para uma análise e interpretação manual (Kantardzic, 2011).

A escolha é um resultado de um processo que envolve avaliação e julgamento; isto é, a avaliação de diferentes opções e tomar uma decisão sobre que opção escolher. De maneira a estes processos acontecerem e a ser feita uma escolha, tem de haver duas ou mais alternativas das quais se possa escolher. Em adição, estas alternativas devem ter algum valor positivo; no sentido que uma “escolha” entre algo que é realmente desejado e algo que não é desejado, não é uma verdadeira escolha (Beresford & Sloper, 2008).

4. Trabalho realizado

Surgiu o interesse de observar melhor a área de venda de veículos novos em Portugal e de tentar perceber como é que as organizações operam no resto do mundo. Como tal, optou-se por questionar informalmente dois gestores de vendas de concessionários no estrangeiro, um da Chevrolet/Ford/Buick na Pensilvânia, EUA que é o maior vendedor do modelo Chevrolet Corvette no nordeste dos Estados Unidos da América e o outro da BMW na Corunha, Espanha, que é o maior vendedor de BMW's da Galiza.

O problema exposto tem a ver com a compra de veículos novos por parte dos concessionários para venda posterior ao público e como podemos reduzir ao máximo esses tempos de venda, para tentar perceber quais são os procedimentos adotados por bons vendedores no estrangeiro.

No caso dos EUA, a informação que se obteve é que tanto a Chevrolet como a Ford disponibilizam um “Modelo de encomenda” com as melhores combinações de motorizações, cores e pacotes de equipamento para a área onde o concessionário está localizado (informação cedida pela marca). Normalmente cerca de 60% dos veículos são encomendados segundo o “Modelo de encomenda” da marca. A outra técnica utilizada, que corresponde a 20% dos veículos encomendados, é o seguimento das especificações de veículos vendidos e encomendados por clientes recentemente. Os restantes 20% são mais aleatórios, são baseados na intuição dos vendedores, algo que consideram que se vai vender (cor ou pacotes específicos) ou apenas veículos com todos os opcionais possíveis.

No caso da BMW em Espanha, só começaram a utilizar o “Modelo de encomenda” da marca há cerca de 3 meses e representa apenas 30% das encomendas. Sendo que a maioria das encomendas são feitas no seguimento das especificações de veículos vendidos e encomendados por clientes recentemente, representam 60% das encomendas. Os restantes 10% são para veículos totalmente equipados que normalmente são mais utilizados para demonstração no concessionário de forma a mostrar ao cliente todas as funcionalidades disponíveis em tempo real.

Onde se denotou mais espaço para melhoramentos foi nas encomendas que são feitas no seguimento das especificações de veículos vendidos e encomendados por clientes recentemente. Ambos os gestores de vendas relataram que fazem uma reunião semanal com os vendedores para saber o que foi pedido por clientes durante aquela semana, em que apenas é trocado verbalmente o que foi pedido, em que no máximo se faz um registo em papel que é descartado após as

encomendas daquela semana, todo o peso das encomendas é colocado no gestor de vendas e na sua memória de vendas passadas.

É de realçar que os gestores de venda, os vendedores, os concessionários e até as marcas têm interesse em vender veículos com mais equipamento pois assim irão obter maiores lucros.

4.1 Dados dos consumidores

O questionário foi elaborado para recolher dados sobre as preferências dos consumidores sobre características de veículos, tais como a gama, cor exterior e interior, caixa de velocidades, combustível, sistema de navegação, ar condicionado, faróis de nevoeiro, bluetooth, internet, sensores de estacionamento, ecrã táctil, jantes, teto de abrir/panorâmico e muitos outros, perfazendo um total de 26 combinações de características dos veículos. Foi previamente realizado um questionário de teste a um grupo de 10 consumidores, em todas as questões havia uma secção para comentários e exposição de dúvidas que ajudou a moldar o questionário até à sua versão final, onde por exemplo se mudou o nome de alguns opcionais, pois estavam a gerar confusão aos inquiridos. Após as retificações, o questionário foi difundido online, de uma forma privada.

O questionário nas perguntas sobre a cor exterior e interior, acabamento dos bancos, caixa de velocidades e combustível disponibiliza um conjunto de respostas possíveis, possibilitando a escolha de uma ou várias respostas. Nas restantes perguntas, as respostas baseiam-se numa escala de um a cinco de forma a perceber a importância dada a cada característica.

Obtiveram-se 100 respostas de consumidores e nas tabelas 1 a 5 podem ser analisadas a caracterização desses consumidores.

Tabela 1 – Género

MASCULINO	FEMININO
52%	48%

Tabela 2 – Idades

18-24	25-34	35-44	45-54	55-64	65+
59%	24%	6%	9%	2%	0%

Tabela 3 - Grau de escolaridade

BÁSICO	SECUNDÁRIO OU PROFISSIONAL	LICENCIATURA	MESTRADO/ DOUTORAMENTO
12%	41%	41%	6%

Tabela 4 – Profissão

EMPR.	DESEMPR.	DOMÉST.	ESTUD.	REFORM.	OUTRO
51%	9%	1%	35%	1%	3%

Tabela 5 - Classificação da situação financeira

PÉSSIMA	MÁ	RAZOÁVEL	BOA	MUITO BOA
2%	11%	54%	31%	2%

Procedeu-se à análise das restantes respostas, para isso, efetuou-se uma operação de transformação nos dados, pois verifica-se que alguns tipos de resposta não foram contemplados pelos consumidores. De seguida, os dados foram divididos por gamas e dentro de cada gama foram analisadas as combinações de cores exteriores e interiores e todos os opcionais presentes. Os opcionais foram estudados em termos percentuais dentro de cada combinação de cores, como tal, foram considerados os que foram iguais ou superiores a 50% como respostas positivas, ou seja, são um requisito obrigatório, os que obtiveram respostas abaixo de 50% foram considerados como não sendo obrigatórios e, dessa forma, tanto podem estar presentes como ausentes. Assim sendo, na tabela 6 é apresentada uma legenda dos opcionais e na tabela 7 o valor que os opcionais 1, 2 e 4 podem obter. Na tabela 8 são apresentadas as combinações de características comuns dentro de cada gama.

Tabela 6 - Listagem dos opcionais sobre que os consumidores foram inquiridos

Nº	OPCIONAL	Nº	OPCIONAL
1	Combustível	14	Ecrã tátil
2	Caixa de velocidades	15	Internet
3	Jantes	16	Espelhos anti encadeamento
4	Acabamento dos bancos	17	Sensor de ângulo morto
5	Pack Acabamentos Exteriores	18	Sistema de som melhorado
6	Sensores de chuva	19	Vidros escurecidos
7	Sensores de luz	20	USB
8	Faróis de nevoeiro	21	Bluetooth
9	Espelhos eletr. reguláveis	22	Controlos no volante
10	Volante/bancos aquecidos	23	Sensores de estacionamento
11	Bancos refrigerados	24	Isofix
12	Bancos eletr. reguláveis	25	Ar condicionado
13	Navegação	26	Teto de abrir e/ou panorâmico

Tabela 7 – Valores dos Opcionais (1), (2) e (4)

COMBUSTÍVEL (1)		CAIXA DE VELOCIDADES (2)		ACABAMENTO DOS BANCOS (4)	
Diesel	“D”	Manual	M	Tecido	T
Gasolina	“G”	Automática	A	Pele	P
Elétrico	“E”	Indiferente	I	Indiferente	I

Na tabela 8, somente são mostradas combinações, por gama e cor exterior e interior. Os valores dos opcionais 1, 2 e 4 estão descritos na tabela 7, os restantes tomam o valor “S” (obrigatório) ou “-” (não obrigatório).

Tabela 8 – Combinações de características - dados do questionário

GAMA	COR EXT.	COR INT.	N °	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Citadino (40%)	Preto	Preto	24	D	M	-	-	-	S	S	S	S	-	S	S	S	S	S	-	S	S	S	S	S	S	S	S	S
	Preto	Outro	9	D	M	-	I	-	S	S	S	S	-	S	-	S	S	S	-	S	S	-	S	S	S	S	S	S
	Azul escuro	Preto	4	G	A	S	T	S	S	S	S	S	-	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Desport./ Cabrio (28%)	Preto	Preto	24	G	M	S	-	-	-	-	S	S	-	S	-	S	S	S	-	S	S	S	S	S	S	S	S	S
Familiar (17%)	Preto	Preto	16	D	M	-	-	-	-	S	S	S	-	S	-	S	S	S	-	S	S	S	S	S	S	S	S	S
SUV/Jipe (12%)	Preto	Preto	8	E	I	-	I	-	-	S	S	S	S	S	S	S	S	S	-	-	S	S	S	S	S	S	S	S
Luxo/Supercarro (3%)	Preto	Preto	3	E	M	-	P	-	-	S	S	S	-	-	-	S	S	S	-	-	-	-	S	-	S	-	S	

Numa breve análise aos resultados obtidos verifica-se que há um padrão nas preferências dos consumidores, principalmente no que diz respeito aos opcionais. Há alguns que são escolhidos em todas as combinações de características, como o ar condicionado (opcional nº 25), faróis de nevoeiro (8), sensores de estacionamento (23), espelhos anti encadeamento (16) e espelhos eletronicamente reguláveis (9). Enquanto há opcionais que são específicos de uma combinação de características, como é o caso das jantes (3), na gama citadino para a combinação de cores exterior e interiores, azul-escuro e preto, respetivamente ou o acabamento dos bancos (4) em pele na gama luxo/supercarro. Verifica-se que ninguém escolheu o opcional teto de abrir e/ou panorâmico (26), por isso não aparece na tabela 8.

4.2 Dados dos concessionários

Foram recolhidos dados reais de concessionários automóveis, de combinações de características de veículos que foram vendidos, bem como o número de dias em que o veículo esteve em stock. Recolheram-se 5.021 registos de vendas realizadas por concessionários de várias marcas em Portugal. Realizou-se um tratamento aos dados recolhidos em que nos veículos vendidos, cada modelo foi inserido numa gama (foram utilizadas as mesmas 5 gamas sobre as quais os consumidores foram inquiridos) e cada cor exterior e interior foi convertida numa cor padrão, facilitando assim a análise.

Analisando os dados, verifica-se que só estão disponíveis 4 opcionais para estudo: 1 - jantes, 2 - caixa de velocidades (se é caixa automática), 3 - navegação e 4- sensores de estacionamento. Os dados foram agrupados por gamas e dentro de cada gama são estudadas as combinações de cores exteriores e interiores e os opcionais relevantes, com o objetivo de perceber que combinações de características de veículos vendem, conseguindo dessa forma analisar as combinações mais pedidas e as que se venderam mais rapidamente, que não são necessariamente as mesmas. Os opcionais foram estudados em termos percentuais dentro de cada combinação de cores; como tal, foram considerados os que foram iguais ou superiores a 50% como combinações positivas, ou seja, é um requisito obrigatório, os restantes opcionais foram considerados como não sendo obrigatórios, ou seja, tanto podem estar presentes como ausentes.

Na tabela 9 são apresentadas as combinações de características comuns dentro de cada gama; a primeira combinação dentro de cada gama (ou as 2 primeiras no caso da gama citadino), representam as combinações mais vendidas em cada gama, a segunda combinação (terceira no caso do citadino), é a combinação que vende mais rapidamente, mas não tem uma grande representação nas vendas. A transformação e posterior análise dos dados resultou nas seguintes combinações de características, o número presente na coluna N° representa o número total de combinações de características e os números presentes nas últimas 3 colunas, MÍN (mínimo), MÁX (máximo) e MÉD (médio) que cada combinação ficou em stock em dias.

Tabela 9 - Combinações de vendas - dados do concessionário

GAMA	COR EXT.	COR INT.	Nº	Op. 1	Op. 2	Op. 3	Op. 4	MÍN	MÁX	MÉD
Citadino (41%)	Preto	Cinzento escuro	73	-	-	-	-	7	432	91
	Branco	Cinzento escuro	73	-	-	-	-	0	562	115
	Laranja	Preto	2	-	-	-	-	41	61	51
Desp./C abrio (6%)	Preto	Preto	18	-	-	-	S	13	630	204
	Verde escuro	Preto	7	-	-	-	-	19	154	56
Familiar (35%)	Preto	Preto	108	-	S	S	S	4	729	203
	Azul claro	Cinzento escuro	1	-	-	S	-	20	20	20
SUV/Jip e (17%)	Preto	Preto	51	-	-	S	-	1	701	138
	Preto	Branco	1	-	-	-	-	45	45	45
Luxo/ Superca rro (1%)	Cinzento escuro	Preto	7	S	-	-	S	57	693	275
	Cinzento escuro	Branco	1	S	S	-	S	17	17	17

4.3 *Data mining*

Depois das análises anteriores que facilitaram a melhor compreensão dos dados, tendo permitido obter alguns resultados, mas visto a sua dimensão foi necessário recorrer a outras heurísticas, baseadas em DM, para prever melhor as configurações que mais vendem. Para tentar obter melhores respostas ao problema em questão, tomou-se por opção procurar modelos que consigam prever melhor as boas (positivas; que pertencem ao(s) intervalo(s) com o número máximo de dias em stock mais perto de zero) e más (negativas; que pertencem ao(s) intervalo(s) com o número máximo de dias em stock mais distante de zero) vendas dentro de cada gama em específico. Depois da análise manual aos dados percebe-se que as diferentes gamas não podem ser tratadas como iguais visto terem opcionais e principalmente tempos de venda diferentes. Para satisfazer tal necessidade, foram criados 15 cenários (Tabela 10), com diferentes combinações de variáveis e 4 Targets (Tabela 11) diferentes, em que o número de dias que cada veículo demorou a ser vendido faz com que este se enquadre numa das classes da Target, também tem o objetivo de conseguir dividir as instâncias em mais subclasses e de ser lógico relativamente à área de negócio onde se insere. Para facilitar a operação do software, os dados sofreram mais algumas transformações. Foram reduzidas as variáveis quer nas cores exteriores, quer nas interiores, os registos com cores com fraca representação, ou foram movidos para a cor “outros” ou então procedeu-se a uma junção de dois ou mais tons da mesma cor numa única, reduzindo o número de cores exteriores para 6 e as interiores para 4, o “nome” de cada cor foi ainda representado apenas por uma letra (tabela 12). Depois de todas as transformações recorreu-se à técnica de mineração de dados, *classification* (classificação), devido ao fato de estarmos a tentar prever uma variável discreta (Data Mining descritivo), neste caso uma das classes dos Targets criados (A; B; C ou D). Para tornar possível esta análise, foi utilizada a ferramenta de software, Weka (Waikato Environment for Knowledge Analysis), principalmente devido a ser open source (uso gratuito) e também devido à disponibilidade de informação sobre a sua utilização. Utilizando a ferramenta, foram explorados os dados de cada Gama para cada Target e Cenário diferente, sendo possível visualizar como output uma Matriz de confusão e ainda outras métricas disponibilizadas pelo software.

Na Tabela 11 o “X” representa a inclusão da variável no cenário e na Tabela 12 cada letra representa um dos intervalos nas diferentes Targets.

Tabela 10 - Combinações de variáveis nos diferentes cenários

Cenário	Cor Exterior	Cor Interior	Jantes	Caixa Automática	Navegação	Sensores Estaciona	Dias Stock
S1	X	X	X	X	X	X	X
S2	X	X					X
S3			X	X	X	X	X
S4		X	X	X	X	X	X
S5	X		X	X	X	X	X
S6	X	X		X	X	X	X
S7	X	X	X		X	X	X
S8	X	X	X	X		X	X
S9	X	X	X	X	X		X
S10	X						X
S11		X					X
S12			X				X
S13				X			X
S14					X		X
S15						X	X

Tabela 11 - Diferentes valores ocupados pelas classes nos diferentes Targets

Targets		T1	T2	T3	T4
Classes	A	0-30	0-9	0-14	0-59
	B	>=31	10-19	15-29	60-179
	C		20-29	30-59	>=180
	D		>=30	>=60	

Tabela 12 – Cores analisadas e a sua representação nos dados

Cores Exteriores	Representação	Cores Interiores	Representação
Azul	a	Preto	p
Branco	b	Cinzeno escuro	r
Preto	p	Castanho	m
Cinzeno escuro	r	Outros	o
Cinzeno claro	s		
Outros	o		

Classificação

Com os dados transformados, os Targets e os Cenários criados, procedeu-se a avaliação dos dados com 3 Algoritmos de classificação, o Naives Bayes, o J48 e o Logistic, utilizando *k-fold cross-validation* com 10 *folds*, onde foram utilizadas a totalidade das instâncias de cada Gama.

O software fornece diversas métricas para avaliar os modelos. A métrica à qual se deu mais importância foi a Acuidade, pois é o número de instâncias que se conseguem prever corretamente dentro de todas as instâncias estudadas, na qual se tentou obter o valor percentual mais elevado, seguido pela sensibilidade e pela especificidade. Também foi considerado importante o ROC (AUC). Para os Targets com 2 classes, utilizamos a acuidade, sensibilidade, especificidade e ROC (AUC), para os Targets com 3 ou classes utilizamos apenas a acuidade e o ROC (AUC).

5. Discussão de resultados

5.1 Resultados da análise manual

5.1.1 Comparação por gamas da quantidade de veículos preferidos pelos consumidores com os vendidos pelos concessionários

Comparando por gamas a quantidade de veículos preferidos pelos consumidores com os vendidos pelos concessionários, em termos percentuais, percebemos que só há grande disparidade nas gamas, desportivo/cabrio e familiar, ver tabela 13.

Tabela 13 – Preferências versus vendas por Gama

GAMA	CONSUMIDORES	CONCESSIONÁRIOS
Citadino	40%	41%
Desport./Cabrio	28%	6%
Familiar	17%	35%
SUV/Jipe	12%	17%
Luxo/Supercarro	3%	1%

Isto pode ser explicado porque os consumidores inquiridos são maioritariamente do sexo masculino, e estão compreendidos na faixa etária dos 18 aos 24 anos, que está de acordo com (Vrkljan & Anaby, 2011) que referem que essa caracterização de inquiridos prefere gamas de veículos mais desportivas. Sendo que a gama familiar, escolhida por pessoas que normalmente são casadas e usam os veículos para propósitos familiares (Baltas & Saridakis, 2013), neste caso os inquiridos não representam essa caracterização, isto é, são uma quantidade reduzida no inquérito realizado, justificando assim a elevada escolha de veículos desportivo/cabrio face ao familiar.

5.1.2 Análise aos opcionais que se podem comparar com os dados do concessionário

Tomou-se por opção analisar apenas os opcionais que se podem comparar com os dados do concessionário, como: 1 - jantes, 2 - caixa de velocidades (se é automática), 3 - navegação e 4 - sensores de estacionamento, apresentados na tabela 11.

Tabela 14 - Dados do consumidor com os opcionais comparáveis

GAMA	COR EXT.	COR INT.	Nº	Op. 1	Op. 2	Op. 3	Op. 4
Citadino (40%)	Preto	Preto	24	-	-	S	S
	Preto	Outro	9	-	-	S	S
	Azul escuro	Preto	4	S	S	S	S
Desport./Cabrio (28%)	Preto	Preto	24	S	-	S	S
Familiar (17%)	Preto	Preto	16	-	-	S	S
SUV/Jipe (12%)	Preto	Preto	8	-	-	S	S
Luxo/Supercarro (3%)	Preto	Preto	3	-	-	-	S

5.1.3 Análise aos dados dos concessionários através das combinações obtidas no questionário

Depois da análise efetuada na tabela 14 aos dados do questionário, aplicaram-se as combinações de características obtidas aos dados fornecidos pelos concessionários; apresentando-se na tabela 15 os resultados.

Tabela 15 - Combinações de características de vendas do concessionário

GAMA	COR EXT	COR INT	Nº	Op. 1	Op. 2	Op. 3	Op. 4	MÍN	MÁX	MÉD
Citadino (41%)	Preto	Preto	16	-	-	S	S	8	681	156
	Preto	Outro	51	-	-	S	S	2	485	118
	Azul escuro	Preto	4	S	S	S	S	10	386	111
Desp./Cabrio (6%)	Preto	Preto	20	S	-	S	S	9	670	115
Familiar (35%)	Preto	Preto	357	-	-	S	S	0	729	80
SUV/Jipe (17%)	Preto	Preto	37	-	-	S	S	1	408	68
Luxo/Supercarro (1%)	Preto	Preto	44	-	-	-	S	13	643	188

5.1.4 Análise aos tempos médios das combinações mais vendidas pelos concessionários

Para apoiar melhor a análise, decidiu-se recolher também o tempo mínimo, máximo e médio de dias que cada gama, independente das combinações de características e o tempo de venda geral do concessionário, ver tabela 16.

Tabela 16 - Análise de vendas por gama

GAMA	Nº DE VEÍCULOS	MÍN	MÁX	MÉD
Citadino	2078	0	851	105
Desport./Cabrio	302	0	936	93
Familiar	1742	0	932	104
SUV/Jipe	829	0	792	85
Luxo/Supercarro	70	1	693	166
Todas as Gamas	5021	0	936	99

Na gama citadino, foram analisadas 3 combinações que representam as cores exteriores e interiores mais comuns com os opcionais preferidos pelos consumidores a essa associados, para primeira combinação (Preto; Preto), o concessionário vendeu 16 veículos numa média de 156 dias, na segunda combinação (Preto; Outros), o concessionário vendeu 51 veículos numa média de 118 dias, na terceira combinação (Azul Escuro; Preto), o concessionário vendeu 4 veículos numa média de 111 dias, perfazendo um total de 71 veículos vendidos numa média de 128 dias, sendo que nas duas primeiras combinações de características, a navegação e os sensores de estacionamento são muito desejados pelos inquiridos e na terceira combinação são todos os opcionais, esta gama, tendo em conta todas as combinações de características, vendeu 2.078 veículos numa média de 105 dias.

Na gama desportivo/cabrio, foi analisada a combinação de cores exteriores e interiores mais comum com os opcionais preferidos pelos consumidores a essa associados, para a combinação (Preto; Preto), o concessionário vendeu 20 veículos numa média de 115 dias, sendo que as jantes, a navegação e os sensores de estacionamento são muito desejados pelos inquiridos, esta gama, tendo em conta todas as combinações de características, vendeu 302 veículos numa média de 93 dias.

Na gama familiar, foi analisada a combinação de cores exteriores e interiores mais comum com os opcionais preferidos pelos consumidores a essa associados, para a combinação (Preto; Preto), o concessionário vendeu 357 veículos numa média de 80 dias, sendo que a navegação e os sensores de estacionamento são muito desejados pelos inquiridos, esta gama, tendo em conta todas as combinações de características, vendeu 1.742 veículos numa média de 104 dias.

Na gama SUV/jipe, foi analisada a combinação de cores Exteriores e Interiores mais comum com os opcionais preferidos pelos consumidores a essa associados, para a combinação (Preto; Preto), o concessionário vendeu 37 veículos numa média de 68 dias, sendo que a navegação e os sensores de estacionamento são muito desejados pelos inquiridos, esta gama, tendo em conta todas as combinações de características, vendeu 829 veículos numa média de 85 dias.

Na gama Luxo/Supercarro, foi analisada a combinação de cores exteriores e interiores mais comum com os opcionais preferidos pelos consumidores a essa associados, para a combinação (Preto; Preto), o concessionário vendeu 44 veículos numa média de 188 dias, sendo que os sensores de estacionamento são muito desejados pelos inquiridos, esta gama, tendo em conta todas as combinações de características, vendeu 70 veículos numa média de 166 dias.

Numa análise geral, as combinações de características pedidas pelos inquiridos, resultam em menores tempos médios de venda nas gamas familiar e SUV/jipe. Maior disparidade foi encontrada nas gamas citadino, desportivo/cabrio e luxo/supercarro em que há uma diferença de 22, 23 e 10 dias respetivamente, em que as combinações de características mais desejadas pelos consumidores demoram mais tempo a ser vendidas do que a média de todas as combinações de características da gama. Pode-se afirmar também que ao longo de todas as gamas os sensores de estacionamento são muito desejados por todos os inquiridos, o que também acontece com a navegação, com a exceção da gama luxo/supercarro.

5.2 Análise com técnicas de *Data mining*

5.2.1 Equilíbrio das classes dos Targets

Como já foi referido em cima, as diferentes gamas têm diferentes tempos médios de venda (tabela 16); para prevenir o desequilíbrio das classes do Target, criaram-se 4 targets diferentes, com objetivo de que pelo menos um Target se adequar a cada gama.

Depois de carregar os dados no Weka foi possível verificar o equilíbrio das classes do Target em cada gama.

Citadino



Figura 3: Equilíbrio das classes nos diferentes Targets para a gama Citadino

Como podemos verificar pela leitura dos 4 gráficos na figura 3, o Target com as classes mais equilibradas é o T1, sendo que T3 ainda apresenta um equilíbrio intermedio. Os targets T2 e T4 apresentam uma classe predominante perante todas as outras, logo há mais desequilíbrio entre classes.

Desportivo/Cabrio

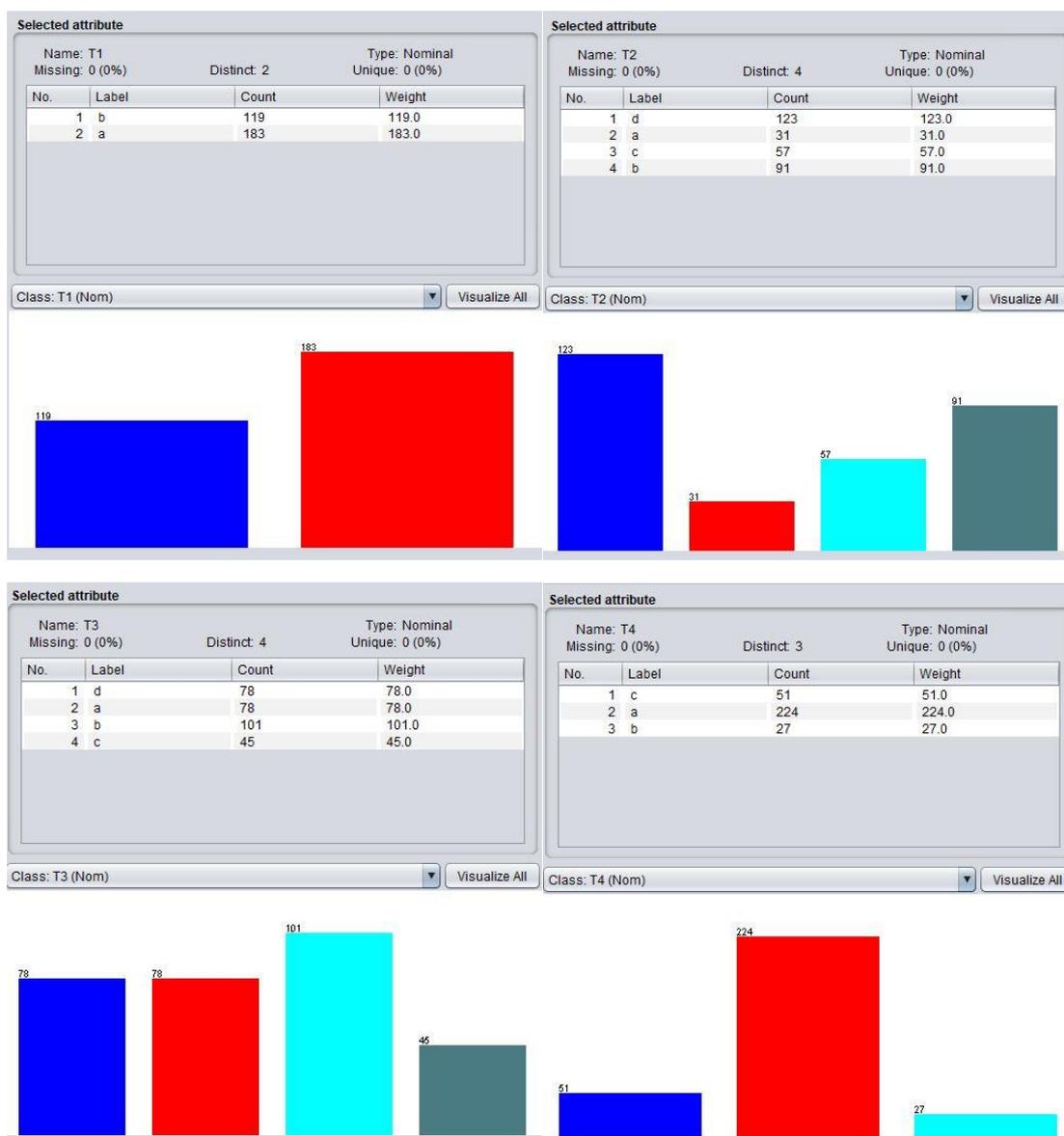


Figura 4: Equilíbrio das classes nos diferentes Targets para a gama Desportivo

Como podemos verificar pela leitura dos 4 gráficos na figura 4, o Target com as classes mais equilibradas é o T1, sendo que T2 e T3 ainda apresentam um equilíbrio intermédio. O target T4 apresenta uma classe predominante perante todas as outras, logo há mais desequilíbrio entre classes.

Familiar



Figura 5: Equilíbrio das classes nos diferentes Targets para a gama Familiar

Como podemos verificar pela leitura dos 4 gráficos na figura 5, o Target com as classes mais equilibradas é o T1, sendo que T3 ainda apresenta um equilíbrio intermédio. Os targets T2 e T4 apresentam uma classe predominante perante todas as outras, logo há mais desequilíbrio entre classes.

Luxo/Supercarro



Figura 6: Equilíbrio das classes nos diferentes Targets para a gama Luxo/Supercarro

Como podemos verificar pela leitura dos 4 gráficos na figura 6, o Target com as classes mais equilibradas é o T1, sendo que T3 ainda apresenta um equilíbrio intermédio. Os targets T2 e T4 apresentam uma classe predominante perante todas as outras, logo há mais desequilíbrio entre classes.

Suv/Jipe



Figura 7: Equilíbrio das classes nos diferentes Targets para a gama SUV/Jipe

Como podemos verificar pela leitura dos 4 gráficos na figura 7, o Target com as classes mais equilibradas é o T1, sendo que T3 ainda apresenta um equilíbrio intermédio. Os targets T2 e T4 apresentam uma classe predominante perante todas as outras, logo há mais desequilíbrio entre classes.

5.2.2 Resultados da Classificação

Nas próximas páginas são apresentados resultados da classificação. Em cada Gama são analisados 15 cenários com 4 Targets, em que cada cenário diferente é analisado com cada Target diferente e com 3 algoritmos diferentes, o Naives Bayes o J48 e o Logistic, em que temos um resultado para cada combinação de cenário, target e algoritmo.

Citadino

A próxima tabela corresponde à análise da Gama Citadino com os algoritmos, Naives Bayes, J48 e Logistic. Realçados a verde estão os que foram considerados resultados mais atrativos para esta Gama. São apresentados apenas os resultados para o target T1, considerados os mais interessantes, os restantes estão disponíveis para consulta em anexo.

Tabela 17 – Resultados da análise aos dados do concessionário na gama citadino para o Target T1

T1	Cenários	Naives Bayes			J48			Logistic		
		Classes	A	B	Avg	A	B	Avg	A	B
		Métricas								
S1	TP Rate	0.538	0.581	0.560	0.641	0.476	0.557	0.590	0.539	0.565
	Precision	0.555	0.564	0.560	0.543	0.577	0.560	0.561	0.569	0.565
	ROC	0.575	0.575	0.575	0.560	0.560	0.560	0.577	0.577	0.577
	Acuidade	0.559			0.557			0.565		
	Sensibilidade	0.580			0.476			0.539		
	Especificidade	0.538			0.641			0.590		
S2	TP Rate	0.502	0.571	0.537	0.536	0.535	0.536	0.512	0.531	0.522
	Precision	0.532	0.541	0.537	0.528	0.543	0.536	0.515	0.528	0.522
	ROC	0.549	0.549	0.549	0.552	0.552	0.552	0.543	0.543	0.543
	Acuidade	0.537			0.536			0.522		
	Sensibilidade	0.571			0.535			0.531		
	Especificidade	0.502			0.536			0.512		
S3	TP Rate	0.578	0.508	0.542	0.499	0.606	0.553	0.605	0.479	0.541
	Precision	0.533	0.553	0.543	0.552	0.555	0.553	0.530	0.556	0.543
	ROC	0.548	0.548	0.548	0.545	0.545	0.545	0.546	0.546	0.546
	Acuidade	0.542			0.553			0.541		
	Sensibilidade	0.508			0.606			0.479		
	Especificidade	0.578			0.499			0.605		
S4	TP Rate	0.488	0.599	0.544	0.599	0.533	0.565	0.498	0.605	0.548
	Precision	0.542	0.542	0.544	0.555	0.578	0.566	0.546	0.550	0.454
	ROC	0.557	0.557	0.557	0.572	0.572	0.572	0.562	0.562	0.562
	Acuidade	0.544			0.565			0.548		
	Sensibilidade	0.599			0.533			0.605		
	Especificidade	0.488			0.599			0.498		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S5	TP Rate	0.541	0.558	0.550	0.574	0.528	0.551	0.548	0.561	0.554
	Precision	0.543	0.556	0.550	0.541	0.560	0.551	0.548	0.561	0.554

	ROC	0.564	0.564	0.564	0.532	0.532	0.532	0.565	0.565	0.565
	Acuidade	0.550			0.551			0.554		
	Sensibilidade	0.558			0.528			0.561		
	Especificidade	0.541			0.574			0.548		
S6	TP Rate	0.452	0.605	0.530	0.596	0.513	0.554	0.479	0.578	0.529
	Precision	0.527	0.532	0.529	0.543	0.566	0.555	0.524	0.533	0.528
	ROC	0.554	0.554	0.544	0.556	0.556	0.556	0.551	0.551	0.551
	Acuidade	0.530			0.554			0.592		
	Sensibilidade	0.605			0.513			0.578		
	Especificidade	0.452			0.596			0.479		
S7	TP Rate	0.521	0.582	0.551	0.585	0.565	0.575	0.494	0.604	0.550
	Precision	0.547	0.555	0.551	0.566	0.583	0.575	0.548	0.552	0.550
	ROC	0.570	0.570	0.570	0.589	0.589	0.589	0.571	0.571	0.571
	Acuidade	0.551			0.575			0.550		
	Sensibilidade	0.582			0.565			0.604		
	Especificidade	0.521			0.585			0.494		
S8	TP Rate	0.544	0.587	0.566	0.630	0.481	0.554	0.534	0.596	0.565
	Precision	0.561	0.570	0.566	0.541	0.572	0.557	0.562	0.568	0.565
	ROC	0.579	0.579	0.579	0.573	0.573	0.573	0.579	0.579	0.579
	Acuidade	0.566			0.554			0.556		
	Sensibilidade	0.587			0.481			0.596		
	Especificidade	0.544			0.630			0.534		
S9	TP Rate	0.544	0.587	0.566	0.667	0.461	0.563	0.532	0.592	0.563
	Precision	0.561	0.570	0.566	0.546	0.588	0.567	0.559	0.566	0.562
	ROC	0.578	0.578	0.578	0.565	0.565	0.565	0.578	0.578	0.578
	Acuidade	0.566			0.563			0.563		
	Sensibilidade	0.587			0.461			0.592		
	Especificidade	0.544			0.667			0.532		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S10	TP Rate	0.245	0.812	0.533	0.245	0.812	0.533	0.262	0.788	0.529
	Precision	0.599	0.525	0.542	0.559	0.525	0.542	0.546	0.524	0.535
	ROC	0.507	0.507	0.507	0.507	0.507	0.507	0.507	0.507	0.507
	Acuidade	0.533			0.533			0.529		
	Sensibilidade	0.812			0.812			0.788		
	Especificidade	0.245			0.245			0.262		
S11	TP Rate	0.565	0.501	0.533	0.565	0.501	0.533	0.565	0.501	0.533
	Precision	0.524	0.543	0.533	0.524	0.543	0.533	0.524	0.543	0.533
	ROC	0.525	0.525	0.525	0.525	0.525	0.525	0.525	0.525	0.525
	Acuidade	0.533			0.533			0.533		
	Sensibilidade	0.501			0.501			0.501		
	Especificidade	0.565			0.565			0.565		
S12	TP Rate	0.686	0.406	0.544	0.686	0.406	0.544	0.686	0.406	0.544
	Precision	0.529	0.571	0.550	0.529	0.571	0.550	0.529	0.571	0.550
	ROC	0.533	0.533	0.533	0.533	0.533	0.533	0.533	0.533	0.533
	Acuidade	0.544			0.544			0.544		
	Sensibilidade	0.406			0.406			0.406		
	Especificidade	0.686			0.686			0.686		
S13	TP Rate	0.688	0.354	0.518	0.688	0.354	0.518	0.688	0.354	0.518
	Precision	0.508	0.538	0.523	0.508	0.538	0.523	0.508	0.538	0.523
	ROC	0.507	0.507	0.507	0.507	0.507	0.507	0.507	0.507	0.507
	Acuidade	0.518			0.518			0.518		
	Sensibilidade	0.354			0.354			0.354		
	Especificidade	0.688			0.688			0.688		
S14	TP Rate	0.461	0.524	0.493	0.074	0.920	0.503	0.461	0.524	0.493
	Precision	0.485	0.500	0.492	0.475	0.506	0.491	0.485	0.500	0.492
	ROC	0.492	0.492	0.492	0.493	0.493	0.493	0.492	0.492	0.492
	Acuidade	0.493			0.503			0.493		
	Sensibilidade	0.524			0.920			0.524		

	Especificidade	0.461			0.074			0.461		
S15	TP Rate	0.055	0.935	0.501	0	1	0.507	0.055	0.935	0.501
	Precision	0.448	0.504	0.477	?	0.507	?	0.448	0.504	0.477
	ROC	0.467	0.467	0.467	0.498	0.498	0.498	0.467	0.467	0.467
	Acuidade	0.501			0.507			0.501		
	Sensibilidade	0.935			1			0.935		
	Especificidade	0.055			0			0.055		

Com os resultados obtidos conseguimos perceber que o desequilíbrio entre as classes dos Targets não nos permitiu ter resultados fidedignos para os Targets T2, T3 e T4, mas o mesmo não se verificou no Target T1, como tal os Targets T2, T3 e T4 são apresentados no anexo I.

Procedemos então com a avaliação dos melhores resultados de cada algoritmo para o Target T1.

Para o algoritmo Naives Bayes, houve 2 cenários que apresentaram os mesmos resultados, o S8 e o S9. Este modelo consegue classificar corretamente 56.6% das instâncias. O rácio de Verdadeiros Positivos e a precisão é quase igual para as duas classes, denota-se que é marginalmente mais elevada para a classe B na qual se consegue classificar corretamente 58.7% das instâncias com 57% de precisão e para a classe A com 54.4% e 56.1% de precisão. A sensibilidade, especificidade e ROC (AUC) são 58.7%, 54.4% e 0.579, respetivamente.

Para o algoritmo J48, o cenário que apresentou os melhores resultados foi o S7. Este modelo consegue classificar corretamente 57.5% das instâncias. O rácio de Verdadeiros Positivos e a precisão é quase igual para as duas classes, denota-se que é marginalmente mais elevada para a classe A na qual se consegue classificar corretamente 58.5% das instâncias com 56.6% de precisão e para a classe B com 56.5% e 58.3% de precisão. A sensibilidade, especificidade e ROC (AUC) são 56.5%, 58.5% e 0.589, respetivamente.

Para o algoritmo Logistic, o cenário que apresentou os melhores resultados foi o S6. Este modelo consegue classificar corretamente 59.2% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe B na qual se consegue classificar corretamente 57.8% das instâncias com 53.3% de precisão e para a classe A com 47.9% e 52.4% de precisão. A sensibilidade, especificidade e ROC (AUC) são 57.8%, 47.9% e 0.551, respetivamente.

Desportivo/Cabrio

As próximas 2 tabelas correspondem à análise da Gama Desportivo/Cabrio com os algoritmos, Naives Bayes, J48 e Logistic. Realçados a verde estão os que foram considerados resultados mais atrativos para esta Gama. São apresentados apenas os resultados para os targets T1 e T3, considerados os mais interessantes, os restantes estão disponíveis para consulta em anexo.

Tabela 18 – Resultados da análise aos dados do concessionário na gama Desportivo/Cabrio para o Target T1

T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S1	TP Rate	0.820	0.252	0.596	0.880	0.076	0.563	0.787	0.311	0.599
	Precision	0.628	0.476	0.568	0.594	0.290	0.608	0.637	0.487	0.578
	ROC	0.588	0.588	0.588	0.461	0.461	0.461	0.589	0.589	0.589
	Acuidade	0.596			0.563			0.599		
	Sensibilidade	0.252			0.076			0.311		
	Especificidade	0.820			0.880			0.787		
S2	TP Rate	0.852	0.193	0.593	1	0	0.606	0.732	0.336	0.576
	Precision	0.619	0.460	0.556	0.606	?	?	0.629	0.449	0.558
	ROC	0.554	0.554	0.554	0.490	0.490	0.490	0.559	0.559	0.559
	Acuidade	0.593			0.606			0.576		
	Sensibilidade	0.193			0			0.336		
	Especificidade	0.852			1			0.732		
S3	TP Rate	0.885	0.134	0.589	0.925	0.025	0.573	0.951	0.059	0.599
	Precision	0.611	0.432	0.541	0.594	0.188	0.434	0.608	0.438	0.541
	ROC	0.546	0.546	0.546	0.493	0.493	0.493	0.544	0.544	0.544
	Acuidade	0.589			0.573			0.599		
	Sensibilidade	0.193			0.025			0.059		
	Especificidade	0.885			0.929			0.951		
S4	TP Rate	0.831	0.227	0.593	0.831	0.210	0.586	0.798	0.277	0.593
	Precision	0.623	0.466	0.561	0.618	0.446	0.550	0.629	0.471	0.567
	ROC	0.569	0.569	0.569	0.499	0.499	0.499	0.566	0.566	0.566
	Acuidade	0.593			0.586			0.593		
	Sensibilidade	0.227			0.210			0.277		
	Especificidade	0.831			0.831			0.798		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S5	TP Rate	0.847	0.176	0.583	0.880	0.109	0.576	0.852	0.193	0.593
	Precision	0.613	0.429	0.540	0.603	0.371	0.512	0.619	0.460	0.593
	ROC	0.571	0.571	0.571	0.521	0.521	0.521	0.565	0.565	0.565
	Acuidade	0.583			0.576			0.593		
	Sensibilidade	0.176			0.109			0.193		
	Especificidade	0.847			0.880			0.852		
S6	TP Rate	0.831	0.218	0.589	0.956	0.008	0.583	0.787	0.311	0.599
	Precision	0.620	0.456	0.556	0.597	0.111	0.406	0.637	0.487	0.578
	ROC	0.590	0.590	0.590	0.463	0.463	0.463	0.589	0.589	0.589
	Acuidade	0.589			0.583			0.599		
	Sensibilidade	0.218			0.008			0.311		
	Especificidade	0.831			0.956			0.787		
S7	TP Rate	0.858	0.227	0.609	0.902	0.076	0.576	0.776	0.319	0.596
	Precision	0.631	0.509	0.583	0.600	0.333	0.495	0.637	0.481	0.575
	ROC	0.564	0.564	0.564	0.480	0.480	0.480	0.574	0.574	0.574

	Acuidade	0.609			0.576			0.596		
	Sensibilidade	0.227			0.076			0.319		
	Especificidade	0.858			0.902			0.776		
S8	TP Rate	0.814	0.252	0.593	0.885	0.067	0.563	0.787	0.311	0.599
	Precision	0.626	0.469	0.564	0.593	0.276	0.468	0.637	0.487	0.578
	ROC	0.591	0.591	0.591	0.467	0.467	0.467	0.595	0.595	0.595
	Acuidade	0.593			0.563			0.599		
	Sensibilidade	0.252			0.067			0.311		
	Especificidade	0.814			0.885			0.787		
S9	TP Rate	0.820	0.244	0.593	0.940	0.042	0.586	0.809	0.277	0.599
	Precision	0.625	0.468	0.563	0.601	0.313	0.488	0.632	0.485	0.574
	ROC	0.593	0.593	0.593	0.479	0.479	0.479	0.597	0.597	0.597
	Acuidade	0.593			0.586			0.599		
	Sensibilidade	0.244			0.042			0.277		
	Especificidade	0.820			0.940			0.809		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S10	TP Rate	1	0	0.606	1	0	0.606	0.989	0	0.599
	Precision	0.606	?	?	0.606	?	?	0.603	0	0.366
	ROC	0.528	0.528	0.528	0.490	0.490	0.490	0.528	0.528	0.528
	Acuidade	0.606			0.606			0.599		
	Sensibilidade	0			0			0		
	Especificidade	1			1			0.989		
S11	TP Rate	1	0	0.606	1	0	0.606	1	0	0.606
	Precision	0.606	?	?	0.606	?	?	0.606	?	?
	ROC	0.508	0.508	0.508	0.490	0.490	0.490	0.505	0.505	0.505
	Acuidade	0.606			0.606			0.606		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		
S12	TP Rate	1	0	0.606	1	0	0.606	1	0	0.606
	Precision	0.606	?	?	0.606	?	?	0.606	?	?
	ROC	0.487	0.487	0.487	0.490	0.490	0.490	0.487	0.487	0.487
	Acuidade	0.606			0.606			0.606		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		
S13	TP Rate	1	0	0.606	1	0	0.606	1	0	0.606
	Precision	0.606	?	?	0.606	?	?	0.606	?	?
	ROC	0.522	0.522	0.522	0.490	0.490	0.490	0.522	0.522	0.522
	Acuidade	0.606			0.606			0.606		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		
S14	TP Rate	1	0	0.606	1	0	0.606	1	0	0.606
	Precision	0.606	?	?	0.606	?	?	0.606	?	?
	ROC	0.479	0.479	0.479	0.490	0.490	0.490	0.479	0.479	0.479
	Acuidade	0.606			0.606			0.606		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		
S15	TP Rate	1	0	0.606	1	0	0.606	1	0	0.606
	Precision	0.606	?	?	0.606	?	?	0.606	?	?
	ROC	0.481	0.481	0.481	0.490	0.490	0.490	0.481	0.481	0.481
	Acuidade	0.606			0.606			0.606		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		

Tabela 19 – Resultados da análise aos dados do concessionário na gama Desportivo/Cabrio para o Target T3

T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S1	TP Rate	0.231	0.485	0.022	0.231	0.285	0.231	0.495	0.067	0.244	0.298	0.205	0.426	0.156	0.282	0.291
	Precision	0.295	0.308	0.100	0.250	0.259	0.295	0.333	0.200	0.250	0.282	0.254	0.309	0.250	0.306	0.285
	ROC	0.508	0.476	0.560	0.529	0.511	0.543	0.458	0.581	0.480	0.504	0.516	0.545	0.584	0.521	0.507
	Acuidade	0.285					0.298					0.291				
S2	TP Rate	0.256	0.604	0	0.167	0.311	0.115	0.832	0	0.051	0.321	0.295	0.436	0	0.231	0.281
	Precision	0.345	0.321	?	0.241	?	0.346	0.326	?	0.222	?	0.196	0.542	?	0.220	?
	ROC	0.527	0.446	0.588	0.511	0.505	0.526	0.456	0.609	0.475	0.501	0.542	0.456	0.579	0.506	0.509
	Acuidade	0.311					0.321					0.281				
S3	TP Rate	0.013	0.723	0	0.244	0.308	0.026	0.663	0.022	0.218	0.288	0.026	0.673	0	0.218	0.288
	Precision	0.067	0.340	?	0.264	?	0.083	0.325	0.143	0.262	0.219	0.074	0.329	?	0.250	?
	ROC	0.407	0.510	0.431	0.519	0.474	0.492	0.449	0.505	0.509	0.484	0.412	0.504	0.453	0.513	0.475
	Acuidade	0.308					0.288					0.288				
S4	TP Rate	0.231	0.564	0	0.231	0.308	0.205	0.564	0.133	0.269	0.331	0.231	0.574	0.067	0.218	0.318
	Precision	0.333	0.328	0	0.247	0.256	0.381	0.329	0.333	0.304	0.337	0.333	0.345	0.231	0.254	0.301
	ROC	0.494	0.476	0.545	0.502	0.498	0.561	0.498	0.599	0.503	0.531	0.492	0.465	0.568	0.494	0.494
	Acuidade	0.308					0.331					0.318				
T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S5	TP Rate	0.179	0.515	0	0.269	0.288	0.192	0.525	0	0.231	0.285	0.154	0.475	0.044	0.295	0.281
	Precision	0.237	0.321	0	0.269	0.238	0.246	0.323	0	0.247	0.235	0.214	0.320	0.100	0.303	0.255
	ROC	0.469	0.492	0.494	0.552	0.502	0.490	0.488	0.533	0.506	0.500	0.470	0.471	0.526	0.547	0.499
	Acuidade	0.288					0.285					0.281				
S6	TP Rate	0.218	0.465	0.044	0.269	0.288	0.192	0.495	0	0.244	0.278	0.205	0.396	0.178	0.333	0.298
	Precision	0.293	0.290	0.035	0.296	0.276	0.263	0.318	0	0.238	0.236	0.267	0.305	0.242	0.333	0.293
	ROC	0.520	0.469	0.574	0.538	0.515	0.498	0.425	0.552	0.492	0.480	0.524	0.451	0.600	0.530	0.513
	Acuidade	0.288					0.278					0.298				

S7	TP Rate	0.256	0.495	0	0.295	0.308	0.205	0.436	0	0.231	0.258	0.192	0.436	0.044	0.333	0.288
	Precision	0.313	0.318	0	0.307	0.266	0.246	0.299	0	0.220	0.220	0.242	0.303	0.133	0.325	0.268
	ROC	0.500	0.470	0.554	0.523	0.504	0.519	0.441	0.542	0.481	0.487	0.516	0.460	0.558	0.523	0.505
	Acuidade	0.308					0.258					0.288				
S8	TP Rate	0.205	0.446	0.044	0.282	0.281	0.256	0.426	0	0.282	0.281	0.205	0.416	0.133	0.282	0.285
	Precision	0.258	0.304	0.143	0.282	0.262	0.253	0.328	0	0.250	0.240	0.254	0.302	0.222	0.301	0.278
	ROC	0.515	0.474	0.566	0.533	0.514	0.481	0.474	0.540	0.421	0.472	0.527	0.462	0.591	0.526	0.514
	Acuidade	0.281					0.281					0.285				
S9	TP Rate	0.231	0.535	0.044	0.231	0.305	0.295	0.455	0.044	0.154	0.275	0.218	0.465	0.133	0.231	0.291
	Precision	0.310	0.320	0.222	0.273	0.291	0.291	0.336	0.095	0.185	0.249	0.279	0.326	0.207	0.265	0.280
	ROC	0.531	0.475	0.581	0.517	0.516	0.572	0.499	0.603	0.449	0.521	0.543	0.454	0.596	0.501	0.510
	Acuidade	0.305					0.275					0.291				
T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0.141	0.644	0	0.064	0.268	0	1	0	0	0.334	0.218	0.554	0	0.077	0.262
	Precision	0.208	0.311	?	0.125	?	?	0.334	?	?	?	0.262	0.296	?	0.125	?
	ROC	0.485	0.464	0.513	0.517	0.490	0.484	0.493	0.468	0.483	0.485	0.486	0.464	0.515	0.518	0.491
	Acuidade	0.268					0.334					0.262				
S11	TP Rate	0.308	0.703	0	0.115	0.344	0.128	0.812	0	0.051	0.318	0.308	0.703	0	0.115	0.344
	Precision	0.393	0.338	?	0.290	?	0.313	0.328	?	0.200	?	0.393	0.338	?	0.290	?
	ROC	0.503	0.488	0.554	0.483	0.487	0.507	0.461	0.601	0.460	0.494	0.502	0.440	0.542	0.482	0.482
	Acuidade	0.344					0.318					0.344				
S12	TP Rate	0	1	0	0	0.334	0	1	0	0	0.334	0	1	0	0	0.334
	Precision	?	0.334	?	?	?	?	0.334	?	?	?	?	0.334	?	?	?
	ROC	0.465	0.487	0.350	0.477	0.458	0.484	0.493	0.468	0.483	0.485	0.465	0.487	0.348	0.477	0.458
	Acuidade	0.334					0.334					0.334				
S13	TP Rate	0	0.832	0	0.115	0.308	0	0.921	0	0.051	0.321	0	0.832	0	0.115	0.308
	Precision	?	0.324	?	0.209	?	?	0.329	?	0.211	?	?	0.324	?	0.209	?
	ROC	0.476	0.483	0.492	0.481	0.482	0.503	0.471	0.484	0.468	0.480	0.476	0.483	0.492	0.482	0.483
	Acuidade	0.308					0.321					0.308				

S14	TP Rate	0.013	0.931	0	0	0.315	0.013	0.931	0	0	0.315	0.013	0.931	0	0	0.315
	Precision	0.071	0.326	?	?	?	0.071	0.326	?	?	?	0.071	0.326	?	?	?
	ROC	0.457	0.483	0.383	0.426	0.447	0.441	0.461	0.446	0.473	0.457	0.457	0.483	0.381	0.426	0.446
	Acuidade	0.315					0.315					0.315				
S15	TP Rate	0	0.743	0	0.231	0.308	0	0.832	0	0.103	0.305	0	0.713	0	0.295	0.315
	Precision	?	0.326	?	0.250	?	?	0.332	?	0.195	?	?	0.706	?	0.261	?
	ROC	0.351	0.466	0.406	0.484	0.432	0.454	0.450	0.500	0.446	0.458	0.351	0.466	0.410	0.484	0.433
	Acuidade	0.308					0.305					0.315				

Com os resultados obtidos conseguimos perceber que o desequilíbrio entre as classes dos Targets não nos permitiu ter resultados fidedignos para os Targets T2 e T4, mas o mesmo não se verificou nos Target T1 e T3, como tal os Targets T2 e T4 são apresentados no anexo II.

Procedemos então com a avaliação dos melhores resultados de cada algoritmo para os Targets T1e T3.

Para o algoritmo Naives Bayes, o cenário que apresentou os melhores resultados foi o S7. Este modelo consegue classificar corretamente 60.9% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe A na qual se consegue classificar corretamente 85.8% das instâncias com 63.1% de precisão e para a classe B com 22.7% e 50.9% de precisão. A sensibilidade, especificidade e ROC (AUC) são 22.7%, 85.8% e 0.564, respetivamente.

Para o algoritmo J48, o cenário que apresentou os melhores resultados foi o S4. Este modelo consegue classificar corretamente 58.6% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe A na qual se consegue classificar corretamente 83.1% das instâncias com 61.8% de precisão e para a classe B com 21% e 44.6% de precisão. A sensibilidade, especificidade e ROC (AUC) são 21%, 83.1% e 0.499, respetivamente.

Para o algoritmo Logistic, houve 3 cenários que apresentaram os mesmos resultados, o S1, o S6 e o S8. Este modelo consegue classificar corretamente 59.9% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é marginalmente mais elevada para a classe A na qual se consegue classificar corretamente 78.7%

das instâncias com 63.7% de precisão e para a classe B com 31.1% e 48.7% de precisão. A sensibilidade, especificidade e ROC (AUC) são 31.1%, 78.7% e 0.595, respetivamente.

O Target T3 apresenta alguns sinais de desequilíbrio entre classes, ainda assim procedemos para a avaliação dos melhores resultados de cada algoritmo.

Para o algoritmo Naives Bayes, o cenário que apresentou os melhores resultados foi o S9. Este modelo consegue classificar corretamente 30.5% das instâncias. O ROC (AUC) do modelo é 0.516, o mais elevado é o da classe C (0.581) seguido pela classe A (0.531), D (0.517) e B (0.475).

Para o algoritmo J48, o cenário que apresentou os melhores resultados foi o S4. Este modelo consegue classificar corretamente 33.1% das instâncias. O ROC (AUC) do modelo é 0.531, o mais elevado é o da classe C (0.599) seguido pela classe A (0.561), D (0.503) e B (0.488).

Para o algoritmo Logistic, o cenário que apresentou os melhores resultados foi o S4. Este modelo consegue classificar corretamente 31.8% das instâncias. O ROC (AUC) do modelo é 0.499, o mais elevado é o da classe C (0.568) seguido pela classe A (0.492), D (0.494) e B (0.465).

Familiar

A próxima tabela corresponde à análise da Gama Familiar com os algoritmos, Naives Bayes, J48 e Logistic. Realçados a verde estão os que foram considerados resultados mais atrativos para esta Gama. São apresentados apenas os resultados para o target T1, considerados os mais interessantes, os restantes estão disponíveis para consulta em anexo.

Tabela 20 – Resultados da análise aos dados do concessionário na gama Familiar para o Target T1

T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S1	TP Rate	0.714	0.363	0.555	0.750	0.335	0.561	0.722	0.351	0.554
	Precision	0.574	0.513	0.546	0.575	0.527	0.553	0.573	0.513	0.545
	ROC	0.572	0.572	0.572	0.543	0.543	0.543	0.572	0.572	0.572
	Acuidade	0.555			0.561			0.554		
	Sensibilidade	0.363			0.335			0.351		
	Especificidade	0.714			0.750			0.722		
S2	TP Rate	0.829	0.187	0.537	1	0	0.546	0.824	0.196	0.539
	Precision	0.551	0.476	0.517	0.546	?	?	0.552	0.481	0.520
	ROC	0.531	0.531	0.531	0.499	0.499	0.499	0.534	0.534	0.534
	Acuidade	0.537			0.546			0.539		
	Sensibilidade	0.187			0			0.196		
	Especificidade	0.829			1			0.824		
S3	TP Rate	0.808	0.288	0.572	0.850	0.235	0.571	0.808	0.288	0.572
	Precision	0.577	0.555	0.567	0.572	0.565	0.569	0.577	0.555	0.567
	ROC	0.557	0.557	0.557	0.543	0.543	0.543	0.555	0.555	0.555

	Acuidade	0.572			0.571			0.572		
	Sensibilidade	0.288			0.235			0.288		
	Especificidade	0.808			0.850			0.808		
S4	TP Rate	0.706	0.372	0.554	0.798	0.321	0.582	0.708	0.378	0.558
	Precision	0.574	0.512	0.546	0.586	0.570	0.578	0.578	0.518	0.551
	ROC	0.566	0.566	0.566	0.569	0.569	0.569	0.567	0.567	0.567
	Acuidade	0.554			0.582			0.558		
	Sensibilidade	0.372			0.321			0.378		
	Especificidade	0.706			0.798			0.708		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S5	TP Rate	0.695	0.353	0.540	0.850	0.236	0.571	0.750	0.288	0.540
	Precision	0.564	0.490	0.530	0.572	0.567	0.570	0.559	0.489	0.527
	ROC	0.568	0.568	0.568	0.540	0.540	0.540	0.564	0.564	0.564
	Acuidade	0.540			0.571			0.540		
	Sensibilidade	0.353			0.236			0.288		
	Especificidade	0.695			0.850			0.750		
S6	TP Rate	0.682	0.388	0.549	0.638	0.391	0.526	0.743	0.295	0.540
	Precision	0.573	0.504	0.542	0.557	0.473	0.519	0.559	0.488	0.527
	ROC	0.565	0.565	0.565	0.520	0.520	0.520	0.562	0.562	0.562
	Acuidade	0.549			0.526			0.540		
	Sensibilidade	0.388			0.391			0.295		
	Especificidade	0.682			0.638			0.743		
S7	TP Rate	0.719	0.346	0.550	0.794	0.278	0.560	0.731	0.359	0.562
	Precision	0.570	0.506	0.541	0.569	0.529	0.551	0.578	0.526	0.554
	ROC	0.567	0.567	0.567	0.553	0.553	0.553	0.568	0.568	0.568
	Acuidade	0.550			0.560			0.562		
	Sensibilidade	0.346			0.278			0.359		
	Especificidade	0.719			0.794			0.731		
S8	TP Rate	0.719	0.350	0.552	0.774	0.312	0.564	0.721	0.353	0.554
	Precision	0.571	0.509	0.543	0.575	0.535	0.557	0.573	0.513	0.545
	ROC	0.563	0.563	0.563	0.518	0.518	0.518	0.564	0.564	0.564
	Acuidade	0.552			0.564			0.554		
	Sensibilidade	0.350			0.312			0.353		
	Especificidade	0.719			0.774			0.721		
S9	TP Rate	0.713	0.369	0.557	0.761	0.316	0.559	0.726	0.355	0.557
	Precision	0.576	0.517	0.549	0.572	0.524	0.550	0.575	0.518	0.549
	ROC	0.575	0.575	0.575	0.527	0.527	0.527	0.574	0.574	0.574
	Acuidade	0.557			0.559			0.557		
	Sensibilidade	0.369			0.316			0.355		
	Especificidade	0.713			0.761			0.726		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S10	TP Rate	0.902	0.073	0.526	1	0	0.546	0.894	0.083	0.526
	Precision	0.539	0.384	0.469	0.546	?	?	0.540	0.395	0.474
	ROC	0.520	0.520	0.520	0.499	0.499	0.499	0.520	0.520	0.520
	Acuidade	0.526			0.546			0.526		
	Sensibilidade	0.073			0			0.083		
	Especificidade	0.902			1			0.894		
S11	TP Rate	0.963	0.027	0.538	1	0	0.546	0.963	0.027	0.538
	Precision	0.543	0.375	0.467	0.546	?	?	0.543	0.375	0.467
	ROC	0.505	0.505	0.505	0.499	0.499	0.499	0.504	0.504	0.504
	Acuidade	0.538			0.546			0.538		
	Sensibilidade	0.027			0			0.027		
	Especificidade	0.963			1			0.963		
S12	TP Rate	0.649	0.437	0.553	0.680	0.391	0.549	0.649	0.437	0.553
	Precision	0.581	0.509	0.548	0.573	0.504	0.542	0.581	0.509	0.548

	ROC	0.529	0.529	0.529	0.529	0.529	0.529	0.529	0.529	0.529
	Acuidade	0.553			0.549			0.553		
	Sensibilidade	0.437			0.391			0.437		
	Especificidade	0.649			0.680			0.649		
S13	TP Rate	1	0	0.546	1	0	0.546	1	0	0.546
	Precision	0.546	?	?	0.546	?	?	0.546	?	?
	ROC	0.515	0.515	0.515	0.499	0.499	0.499	0.515	0.515	0.515
	Acuidade	0.546			0.546			0.546		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		
S14	TP Rate	1	0	0.546	1	0	0.546	1	0	0.546
	Precision	0.546	?	?	0.546	?	?	0.546	?	?
	ROC	0.523	0.523	0.523	0.499	0.499	0.499	0.523	0.523	0.523
	Acuidade	0.546			0.546			0.546		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		
S15	TP Rate	1	0	0.546	1	0	0.546	1	0	0.546
	Precision	0.546	?	?	0.546	?	?	0.546	?	?
	ROC	0.495	0.495	0.495	0.499	0.499	0.499	0.495	0.495	0.495
	Acuidade	0.546			0.546			0.546		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		

Com os resultados obtidos conseguimos perceber que o desequilíbrio entre as classes dos Targets não nos permitiu ter resultados fidedignos para os Targets T2, T3 e T4, mas o mesmo não se verificou no Target T1, como tal os Targets T2, T3 e T4 são apresentados no anexo III.

Procedemos então com a avaliação dos melhores resultados de cada algoritmo para o Target T1.

Para o algoritmo Naives Bayes, o cenário que apresentou os melhores resultados foi o S3. Este modelo consegue classificar corretamente 57.2% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe A na qual se consegue classificar corretamente 80.8% das instâncias com 57.7% de precisão e para a classe B com 28.8% e 55.5% de precisão. A sensibilidade, especificidade e ROC (AUC) são 22.8%, 80.8% e 0.557, respetivamente.

Para o algoritmo J48, o cenário que apresentou os melhores resultados foi o S4. Este modelo consegue classificar corretamente 58.2% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe A na qual se consegue classificar corretamente 79.8% das instâncias com 58.6% de precisão e para a classe B com 32.1% e 57% de precisão. A sensibilidade, especificidade e ROC (AUC) são 32.1%, 79.8% e 0.569, respetivamente.

Para o algoritmo Logistic, o cenário que apresentou os melhores resultados foi o S7. Este modelo consegue classificar corretamente 56.2% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe A na qual se consegue classificar corretamente 73.1% das instâncias com 57.8% de precisão e para a classe B com 35.9% e 52.6% de precisão. A sensibilidade, especificidade e ROC (AUC) são 35.9%, 73.1% e 0.568, respetivamente.

Luxo/Supercarro

As próximas 2 tabelas correspondem à análise da Gama Luxo/Supercarro com os algoritmos, Naives Bayes, J48 e Logistic. Realçados a verde estão os que foram considerados resultados mais atrativos para esta Gama. São apresentados apenas os resultados para os targets T1 e T4, considerados os mais interessantes, os restantes estão disponíveis para consulta em anexo.

Tabela 21 – Resultados da análise aos dados do concessionário na gama Luxo/Supercarro para o Target T1

T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S1	TP Rate	0.407	0.677	0.629	0.185	0.674	0.486	0.407	0.791	0.643
	Precision	0.524	0.673	0.616	0.263	0.569	0.451	0.550	0.680	0.630
	ROC	0.621	0.621	0.621	0.398	0.398	0.398	0.635	0.635	0.635
	Acuidade	0.629			0.486			0.643		
	Sensibilidade	0.677			0.674			0.791		
	Especificidade	0.407			0.185			0.407		
S2	TP Rate	0.074	0.837	0.543	0	1	0.614	0	0.791	0.486
	Precision	0.222	0.590	0.448	?	0.614	?	0	0.557	0.342
	ROC	0.488	0.487	0.487	0.437	0.437	0.437	0.462	0.462	0.462
	Acuidade	0.543			0.614			0.486		
	Sensibilidade	0.837			1			0.791		
	Especificidade	0.074			0			0		
S3	TP Rate	0.333	0.721	0.571	0.259	0.698	0.529	0.296	0.744	0.571
	Precision	0.429	0.633	0.554	0.350	0.600	0.504	0.421	0.627	0.548
	ROC	0.617	0.617	0.617	0.458	0.458	0.458	0.612	0.612	0.612
	Acuidade	0.571			0.529			0.571		
	Sensibilidade	0.721			0.698			0.744		
	Especificidade	0.333			0.259			0.296		
S4	TP Rate	0.296	0.767	0.586	0.259	0.698	0.529	0.444	0.698	0.600
	Precision	0.444	0.635	0.561	0.350	0.600	0.504	0.480	0.667	0.595
	ROC	0.604	0.606	0.605	0.458	0.458	0.458	0.625	0.625	0.625
	Acuidade	0.586			0.529			0.600		
	Sensibilidade	0.767			0.698			0.698		
	Especificidade	0.296			0.259			0.444		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S5	TP Rate	0.444	0.767	0.643	0.185	0.698	0.500	0.481	0.814	0.686
	Precision	0.545	0.688	0.633	0.278	0.577	0.462	0.619	0.714	0.678
	ROC	0.645	0.645	0.645	0.433	0.433	0.433	0.636	0.636	0.636

	Acuidade	0.643			0.500			0.686		
	Sensibilidade	0.767			0.698			0.814		
	Especificidade	0.444			0.185			0.481		
S6	TP Rate	0.296	0.814	0.614	0.148	0.791	0.543	0.333	0.744	0.586
	Precision	0.500	0.648	0.591	0.308	0.596	0.485	0.450	0.640	0.567
	ROC	0.612	0.612	0.612	0.422	0.422	0.422	0.624	0.624	0.624
	Acuidade	0.614			0.543			0.586		
	Sensibilidade	0.814			0.791			0.744		
	Especificidade	0.296			0.148			0.333		
S7	TP Rate	0.296	0.837	0.629	0.185	0.698	0.500	0.370	0.791	0.629
	Precision	0.533	0.655	0.608	0.278	0.577	0.462	0.526	0.667	0.613
	ROC	0.612	0.612	0.612	0.408	0.408	0.408	0.612	0.612	0.612
	Acuidade	0.629			0.500			0.629		
	Sensibilidade	0.837			0.698			0.791		
	Especificidade	0.296			0.185			0.370		
S8	TP Rate	0.407	0.767	0.629	0.185	0.674	0.486	0.407	0.791	0.643
	Precision	0.524	0.673	0.616	0.263	0.569	0.451	0.550	0.680	0.630
	ROC	0.621	0.621	0.621	0.398	0.398	0.398	0.635	0.635	0.635
	Acuidade	0.629			0.486			0.643		
	Sensibilidade	0.767			0.674			0.791		
	Especificidade	0.407			0.185			0.407		
S9	TP Rate	0.296	0.884	0.657	0	0.860	0.529	0.333	0.767	0.600
	Precision	0.615	0.667	0.647	0	0.578	0.355	0.474	0.647	0.580
	ROC	0.554	0.554	0.554	0.366	0.366	0.366	0.554	0.554	0.554
	Acuidade	0.657			0.529			0.600		
	Sensibilidade	0.884			0.860			0.767		
	Especificidade	0.296			0			0.333		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S10	TP Rate	0	0.907	0.557	0	1	0.614	0.074	0.860	0.557
	Precision	0	0.591	0.363	?	0.614	?	0.250	0.597	0.463
	ROC	0.504	0.505	0.505	0.437	0.437	0.437	0.473	0.473	0.473
	Acuidade	0.557			0.614			0.557		
	Sensibilidade	0.907			1			0.860		
	Especificidade	0			0			0.074		
S11	TP Rate	0	0.930	0.571	0	1	0.614	0	0.930	0.571
	Precision	0	0.597	0.367	?	0.614	?	0	0.597	0.367
	ROC	0.342	0.339	0.340	0.437	0.437	0.437	0.383	0.383	0.383
	Acuidade	0.571			0.614			0.571		
	Sensibilidade	0.930			1			0.930		
	Especificidade	0			0			0		
S12	TP Rate	0.185	0.837	0.586	0.111	0.837	0.557	0.185	0.837	0.586
	Precision	0.417	0.621	0.542	0.300	0.600	0.484	0.417	0.621	0.542
	ROC	0.505	0.505	0.505	0.474	0.474	0.474	0.505	0.505	0.505
	Acuidade	0.586			0.557			0.586		
	Sensibilidade	0.837			0.837			0.837		
	Especificidade	0.185			0.111			0.185		
S13	TP Rate	0	1	0.614	0	1	0.614	0	1	0.614
	Precision	?	0.614	?	?	0.614	?	?	0.614	?
	ROC	0.479	0.479	0.479	0.437	0.437	0.437	0.479	0.479	0.479
	Acuidade	0.614			0.614			0.614		
	Sensibilidade	1			1			1		
	Especificidade	0			0			0		
S14	TP Rate	0	1	0.614	0	1	0.614	0	1	0.614
	Precision	?	0.614	?	?	0.614	?	?	0.614	?
	ROC	0.437	0.437	0.437	0.437	0.437	0.437	0.437	0.437	0.437
	Acuidade	0.614			0.614			0.614		
	Sensibilidade	1			1			1		
	Especificidade	0			0			0		

S15	TP Rate	0.407	0.744	0.614	0.259	0.767	0.571	0.519	0.721	0.643
	Precision	0.500	0.667	0.602	0.412	0.623	0.541	0.538	0.705	0.640
	ROC	0.530	0.530	0.530	0.493	0.493	0.493	0.530	0.530	0.530
	Acuidade	0.614			0.571			0.643		
	Sensibilidade	0.744			0.767			0.721		
Especificidade	0.407			0.259			0.519			

Tabela 22 – Resultados da análise aos dados do concessionário na gama Luxo/Supercarro para o Target T4

T4		Naives Bayes				J48				Logistic			
Cenários	Classes	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
S1	TP Rate	0.706	0.250	0.333	0.500	0.912	0	0	0.443	0.676	0.250	0.375	0.500
	Precision	0.558	0.500	0.381	0.487	0.470	0	0	0.228	0.561	0.500	0.391	0.492
	ROC	0.574	0.603	0.502	0.554	0.442	0.397	0.401	0.420	0.610	0.462	0.550	0.564
	Acuidade	0.500				0.434				0.500			
S2	TP Rate	0.676	0	0.417	0.471	0.941	0	0.042	0.471	0.706	0	0.458	0.500
	Precision	0.535	?	0.370	?	0.485	?	0.250	?	0.600	0	0.379	0.421
	ROC	0.592	0.501	0.537	0.557	0.438	0.458	0.410	0.432	0.617	0.421	0.568	0.567
	Acuidade	0.471				0.471				0.500			
S3	TP Rate	0.735	0.417	0	0.429	0.853	0	0	0.414	0.765	0.333	0	0.429
	Precision	0.490	0.357	0	0.299	0.453	0	0	0.220	0.722	0.121	0.152	0.305
	ROC	0.491	0.633	0.385	0.479	0.360	0.471	0.407	0.395	0.492	0.629	0.381	0.478
	Acuidade	0.429				0.414				0.429			
S4	TP Rate	0.765	0.417	0.083	0.471	0.882	0	0	0.429	0.765	0.333	0.083	0.457
	Precision	0.491	0.556	0.250	0.419	0.492	0	0	0.239	0.500	0.444	0.222	0.395
	ROC	0.496	0.636	0.442	0.501	0.477	0.408	0.301	0.405	0.491	0.608	0.446	0.496
	Acuidade	0.471				0.429				0.457			
T4		Naives Bayes				J48				Logistic			
Cenários	Classes	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
S5	TP Rate	0.618	0.167	0.292	0.429	0.912	0	0	0.443	0.647	0.250	0.292	0.457
	Precision	0.525	0.286	0.304	0.408	0.470	0	?	?	0.564	0.375	0.304	0.443
	ROC	0.579	0.597	0.489	0.551	0.443	0.397	0.441	0.434	0.629	0.469	0.503	0.559
	Acuidade	0.429				0.434				0.457			
S6	TP Rate	0.706	0.167	0.375	0.500	0.912	0	0	0.443	0.735	0.250	0.375	0.529

	Precisión	0.533	0.500	0.429	0.492	0.470	0	?	?	0.581	0.429	0.450	0.510
	ROC	0.602	0.606	0.518	0.574	0.442	0.397	0.401	0.420	0.618	0.506	0.560	0.579
	Acuidad e	0.500				0.434				0.529			
	S7	TP Rate	0.765	0	0.375	0.500	1	0	0	0.486	0.706	0	0.417
	Precisión	0.553	0	0.409	0.409	0.486	?	?	?	0.545	0	0.400	0.402
	ROC	0.591	0.522	0.554	0.566	0.431	0.420	0.424	0.427	0.624	0.422	0.590	0.578
	Acuidad e	0.500				0.486				0.486			
	S8	TP Rate	0.706	0.250	0.333	0.500	0.912	0	0	0.443	0.676	0.250	0.375
	Precisión	0.558	0.500	0.381	0.487	0.470	0	0	0.228	0.561	0.500	0.391	0.492
	ROC	0.574	0.603	0.502	0.554	0.442	0.397	0.401	0.420	0.610	0.462	0.550	0.564
	Acuidad e	0.500				0.434				0.500			
	S9	TP Rate	0.618	0.333	0.333	0.471	0.912	0	0	0.443	0.559	0.250	0.375
	Precisión	0.512	0.667	0.348	0.482	0.470	0	0	0.228	0.514	0.333	0.375	0.435
	ROC	0.560	0.606	0.476	0.539	0.442	0.397	0.401	0.420	0.574	0.464	0.514	0.535
	Acuidad e	0.471				0.434				0.434			
	T4		Naives Bayes				J48				Logistic		
Cenários	Classes	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
S10	TP Rate	0.735	0	0.542	0.543	0.941	0	0.042	0.471	0.676	0	0.542	0.514
	Precisión	0.610	?	0.448	?	0.485	?	0.250	?	0.590	?	0.419	?
	ROC	0.590	0.386	0.517	0.530	0.438	0.458	0.410	0.432	0.594	0.431	0.538	0.547
	Acuidad e	0.543				0.471				0.514			
S11	TP Rate	1	0	0	0.486	1	0	0	0.486	1	0	0.042	0.500
	Precisión	0.486	?	?	?	0.486	?	?	?	0.500	?	0.500	?
	ROC	0.428	0.499	0.431	0.441	0.431	0.420	0.424	0.427	0.400	0.439	0.428	0.416
	Acuidad e	0.486				0.486				0.500			
S12	TP Rate	1	0	0	0.486	1	0	0	0.486	1	0	0	0.486
	Precisión	0.486	?	?	?	0.486	?	?	?	0.486	?	?	?
	ROC	0.398	0.522	0.279	0.378	0.431	0.420	0.424	0.427	0.426	0.522	0.337	0.412
	Acuidad e	0.486				0.486				0.486			
S13	TP Rate	0.912	0	0	0.443	0.941	0	0	0.457	0.882	0.083	0	0.443

	Precisão	0.470	0	?	?	0.471	0	?	?	0.484	0.125	?	?
	ROC	0.490	0.550	0.386	0.464	0.409	0.407	0.455	0.424	0.489	0.550	0.397	0.468
	Acuidade	0.434				0.457				0.434			
	TP Rate	1	0	0	0.486	1	0	0	0.486	1	0	0	0.486
S14	Precisão	0.486	?	?	?	0.486	?	?	?	0.486	?	?	?
	ROC	0.431	0.420	0.424	0.427	0.431	0.420	0.424	0.427	0.431	0.420	0.424	0.427
	Acuidade	0.486				0.486				0.486			
	TP Rate	1	0	0	0.486	1	0	0	0.486	0.824	0	0.125	0.443
S15	Precisão	0.486	?	?	?	0.486	?	?	?	0.467	?	0.300	?
	ROC	0.485	0.453	0.476	0.477	0.431	0.420	0.424	0.427	0.485	0.453	0.476	0.477
	Acuidade	0.486				0.486				0.434			
	TP Rate	1	0	0	0.486	1	0	0	0.486	0.824	0	0.125	0.443

Com os resultados obtidos conseguimos perceber que o desequilíbrio entre as classes dos Targets não nos permitiu ter resultados fidedignos para os Targets T2 e T3, mas o mesmo não se verificou nos Target T1 e T4, como tal os Targets T2 e T3 são apresentados no anexo IV.

Procedemos então com a avaliação dos melhores resultados de cada algoritmo para os Targets T1 e T4.

Para o algoritmo Naives Bayes, o cenário que apresentou os melhores resultados foi o S9. Este modelo consegue classificar corretamente 65.7% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe B na qual se consegue classificar corretamente 88.4% das instâncias com 66.7% de precisão e para a classe A com 29.6% e 61.5% de precisão. A sensibilidade, especificidade e ROC (AUC) são 88.4%, 29.6% e 0.554, respetivamente.

Para o algoritmo J48, o cenário que apresentou os melhores resultados foi o S15. Este modelo consegue classificar corretamente 57.1% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe B na qual se consegue classificar corretamente 76.6% das instâncias com 62.3% de precisão e para a classe A com 29.5% e 41.2% de precisão. A sensibilidade, especificidade e ROC (AUC) são 77.6%, 25.9% e 0.493, respetivamente.

Para o algoritmo Logistic, o cenário que apresentou os melhores resultados foi o S5. Este modelo consegue classificar corretamente 68.6% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe B na qual se consegue classificar corretamente 81.4% das instâncias com 71.4% de precisão e para a classe A com 48.1% e 61.9% de precisão. A sensibilidade, especificidade e ROC (AUC) são 81.4%, 48.1% e 0.636, respetivamente.

O Target T4 apresenta alguns sinais de desequilíbrio entre classes, ainda assim procedemos para a avaliação dos melhores resultados de cada algoritmo.

Para o algoritmo Naives Bayes, o cenário que apresentou os melhores resultados foi o S6. Este modelo consegue classificar corretamente 50% das instâncias. O ROC (AUC) do modelo é 0.574, o mais elevado é o da classe B (0.606) seguido pela classe A (0.602) e B (0.518).

Para o algoritmo J48, não foi possível encontrar um modelo que não tivesse sofrido com o desequilíbrio das classes do Target.

Para o algoritmo Logistic, o cenário que apresentou os melhores resultados foi o S6. Este modelo consegue classificar corretamente 52.9% das instâncias. O ROC (AUC) do modelo é 0.579, o mais elevado é o da classe A (0.618) seguido pelas classes B (0.506) e C (0.506).

SUV/Jipe

A próxima tabela corresponde à análise da Gama SUV/Jipe com os algoritmos, Naives Bayes, J48 e Logistic. Realçados a verde estão os que foram considerados resultados mais atrativos para esta Gama. São apresentados apenas os resultados para o target T1, considerados os mais interessantes, os restantes estão disponíveis para consulta em anexo.

Tabela 23 – Resultados da análise aos dados do concessionário na gama SUV/Jipe para o Target T1

T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S1	TP Rate	0.700	0.361	0.537	0.595	0.544	0.571	0.681	0.348	0.521
	Precision	0.541	0.527	0.535	0.584	0.555	0.570	0.530	0.504	0.517
	ROC	0.517	0.517	0.517	0.559	0.559	0.559	0.519	0.519	0.519
	Acuidade	0.537			0.571			0.521		
	Sensibilidade	0.361			0.544			0.348		
	Especificidade	0.700			0.595			0.681		
S2	TP Rate	0.760	0.308	0.543	0.805	0.276	0.550	0.760	0.308	0.543
	Precision	0.542	0.554	0.543	0.545	0.567	0.556	0.542	0.544	0.543
	ROC	0.514	0.514	0.514	0.503	0.503	0.503	0.509	0.509	0.509

	Acuidade	0.543			0.550			0.543		
	Sensibilidade	0.308			0.276			0.308		
	Especificidade	0.760			0.805			0.760		
S3	TP Rate	0.800	0.198	0.510	0.560	0.551	0.556	0.786	0.228	0.517
	Precision	0.518	0.479	0.499	0.574	0.538	0.557	0.523	0.497	0.511
	ROC	0.502	0.502	0.502	0.567	0.567	0.567	0.505	0.505	0.505
	Acuidade	0.510			0.556			0.517		
	Sensibilidade	0.198			0.551			0.228		
	Especificidade	0.800			0.560			0.786		
S4	TP Rate	0.721	0.241	0.538	0.567	0.586	0.577	0.709	0.353	0.538
	Precision	0.541	0.531	0.536	0.597	0.557	0.578	0.542	0.530	0.536
	ROC	0.531	0.531	0.531	0.590	0.590	0.590	0.531	0.531	0.531
	Acuidade	0.538			0.577			0.538		
	Sensibilidade	0.241			0.586			0.335		
	Especificidade	0.721			0.567			0.709		
S5	TP Rate	0.658	0.338	0.504	0.551	0.549	0.550	0.656	0.323	0.496
	Precision	0.517	0.479	0.499	0.568	0.532	0.551	0.511	0.466	0.489
	ROC	0.491	0.491	0.491	0.558	0.558	0.558	0.493	0.493	0.493
	Acuidade	0.504			0.550			0.496		
	Sensibilidade	0.338			0.549			0.323		
	Especificidade	0.658			0.551			0.656		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S6	TP Rate	0.702	0.338	0.527	0.630	0.536	0.585	0.670	0.353	0.517
	Precision	0.534	0.513	0.524	0.594	0.574	0.584	0.527	0.498	0.513
	ROC	0.507	0.507	0.507	0.567	0.567	0.567	0.508	0.508	0.508
	Acuidade	0.527			0.585			0.517		
	Sensibilidade	0.338			0.536			0.335		
	Especificidade	0.702			0.630			0.670		
S7	TP Rate	0.677	0.376	0.529	0.700	0.366	0.539	0.677	0.371	0.530
	Precision	0.539	0.519	0.529	0.543	0.531	0.537	0.537	0.516	0.527
	ROC	0.513	0.513	0.513	0.550	0.550	0.550	0.513	0.513	0.513
	Acuidade	0.529			0.539			0.530		
	Sensibilidade	0.376			0.366			0.371		
	Especificidade	0.677			0.700			0.667		
S8	TP Rate	0.707	0.348	0.534	0.651	0.439	0.549	0.695	0.343	0.526
	Precision	0.539	0.525	0.532	0.556	0.538	0.547	0.533	0.511	0.522
	ROC	0.519	0.519	0.519	0.536	0.536	0.536	0.520	0.520	0.520
	Acuidade	0.534			0.549			0.526		
	Sensibilidade	0.348			0.439			0.334		
	Especificidade	0.707			0.651			0.695		
S9	TP Rate	0.705	0.358	0.538	0.579	0.556	0.568	0.705	0.343	0.531
	Precision	0.542	0.530	0.536	0.585	0.551	0.568	0.536	0.519	0.528
	ROC	0.528	0.528	0.528	0.569	0.569	0.569	0.530	0.530	0.530
	Acuidade	0.538			0.568			0.531		
	Sensibilidade	0.358			0.556			0.334		
	Especificidade	0.705			0.579			0.705		
T1		Naives Bayes			J48			Logistic		
Cenários	Classes	A	B	Avg	A	B	Avg	A	B	Avg
	Métricas									
S10	TP Rate	0.691	0.288	0.497	0.900	0.080	0.505	0.672	0.313	0.499
	Precision	0.511	0.464	0.488	0.513	0.427	0.472	0.513	0.470	0.492
	ROC	0.475	0.475	0.475	0.478	0.478	0.478	0.474	0.474	0.474
	Acuidade	0.497			0.505			0.499		
	Sensibilidade	0.288			0.080			0.331		
	Especificidade	0.691			0.900			0.672		
S11	TP Rate	0.802	0.283	0.552	0.802	0.283	0.552	0.802	0.283	0.552
	Precision	0.547	0.571	0.558	0.547	0.571	0.558	0.547	0.571	0.558

	ROC	0.506	0.506	0.506	0.506	0.506	0.506	0.506	0.506	0.506
	Acuidade	0.552			0.552			0.552		
	Sensibilidade	0.283			0.283			0.283		
	Especificidade	0.802			0.802			0.802		
S12	TP Rate	0.772	0.271	0.531	0.772	0.271	0.531	0.772	0.271	0.531
	Precision	0.533	0.524	0.529	0.533	0.524	0.529	0.533	0.524	0.529
	ROC	0.502	0.502	0.502	0.502	0.502	0.502	0.502	0.502	0.502
	Acuidade	0.531			0.531			0.531		
	Sensibilidade	0.271			0.271			0.271		
	Especificidade	0.772			0.772			0.772		
S13	TP Rate	1	0	0.519	1	0	0.519	0.940	0.050	0.511
	Precision	0.519	?	?	0.519	?	?	0.516	0.435	0.477
	ROC	0.493	0.493	0.493	0.499	0.499	0.499	0.493	0.493	0.493
	Acuidade	0.519			0.519			0.51		
	Sensibilidade	0			0			0.050		
	Especificidade	1			1			0.940		
S14	TP Rate	1	0	0.519	1	0	0.519	1	0	0.519
	Precision	0.519	?	?	0.519	?	?	0.519	?	?
	ROC	0.485	0.485	0.485	0.499	0.499	0.499	0.485	0.485	0.485
	Acuidade	0.519			0.519			0.519		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		
S15	TP Rate	1	0	0.519	1	0	0.519	1	0	0.519
	Precision	0.519	?	?	0.519	?	?	0.519	?	?
	ROC	0.423	0.423	0.423	0.499	0.499	0.499	0.423	0.423	0.423
	Acuidade	0.519			0.519			0.519		
	Sensibilidade	0			0			0		
	Especificidade	1			1			1		

Com os resultados obtidos conseguimos perceber que o desequilíbrio entre as classes dos Targets não nos permitiu ter resultados fidedignos para os Targets T2, T3 e T4, mas o mesmo não se verificou no Target T1, como tal os Targets T2, T3 e T4 são apresentados no anexo V.

Procedemos então com a avaliação dos melhores resultados de cada algoritmo para o Target T1.

Para o algoritmo Naives Bayes, o cenário que apresentou os melhores resultados foi o S11. Este modelo consegue classificar corretamente 55.2% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe A na qual se consegue classificar corretamente 80.2% das instâncias com 54.7% de precisão e para a classe B com 28.3% e 57.1% de precisão. A sensibilidade, especificidade e ROC (AUC) são 28.3%, 80.2% e 0.506, respetivamente.

Para o algoritmo J48, o cenário que apresentou os melhores resultados foi o S6. Este modelo consegue classificar corretamente 55.8% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe A na qual se consegue classificar corretamente 63% das instâncias com 59.4% de precisão e para a classe

B com 53.6% e 57.4% de precisão. A sensibilidade, especificidade e ROC (AUC) são 53.6%, 63% e 0.567, respetivamente.

Para o algoritmo Logistic, o cenário que apresentou os melhores resultados foi o S2. Este modelo consegue classificar corretamente 54.3% das instâncias. O rácio de Verdadeiros Positivos e a precisão é diferente para as duas classes, denota-se que é mais elevada para a classe A na qual se consegue classificar corretamente 76% das instâncias com 54.2% de precisão e para a classe B com 30.8% e 54.4% de precisão. A sensibilidade, especificidade e ROC (AUC) são 30.8%, 76% e 0.509, respetivamente.

5.2.3 Modelo que obteve melhores resultados para cada Gama

Como já foi apontado, a métrica que será de maior relevância é a acuidade que é a percentagem de instâncias que o modelo consegue classificar corretamente.

Sendo esta a métrica mais relevante, nas Gamas Desportivo e Luxo, os resultados dos Targets T3 e T4 respetivamente, apresentaram valores inferiores de acuidade do que os do Target T1. Aqui concluímos que a criação dos Targets T2, T3 e T4 não funcionaram como era esperado, depois da leitura dos tempos médios e máximos dessas mesmas gamas, considerou-se ser vantajoso criar Targets que considerassem positivas as vendas de mais do que um intervalo, tendo esses intervalos valores positivos mais elevados do que o Target T1 que é 30 dias.

Na Gama Citadino, o algoritmo que apresentou melhores resultados foi o Logistic no cenário S6 (todas as variáveis menos as jantes), com 59.2% de acuidade. A proporção de vendas positivas identificadas pelo teste com uma percentagem de 57.8% (sensibilidade) e de vendas negativas de 47.9% (especificidade). E capacidade preditiva do modelo classificar as vendas positivas e as negativas é 55.1% (ROC=0.551). O facto do modelo que apresentou melhores resultados, tendo em conta as métricas estudadas, ter acontecido no cenário S6, em que estão todas as variáveis presentes menos as Jantes, pode indicar que este opcional é menos importante para prever esta Gama e que por sua vez os consumidores desta Gama estão menos interessados nele, raciocínio que pode ser apoiado pela tabela 14, mostra os resultados simplificados das escolhas dos consumidores no questionário, que também indica que opcional Jantes não é muito desejado pelos inquiridos que têm preferência por esta gama.

Na Gama Desportivo, o algoritmo que apresentou melhores resultados foi o Logistic no cenário S8 (todas as variáveis menos a Navegação), com 59.9% de acuidade. A proporção de Vendas

positivas identificadas pelo teste com uma percentagem de 31.1% (sensibilidade) e de vendas negativas de 78.7% (especificidade). E capacidade preditiva do modelo classificar as vendas positivas e as negativas é 59.5% (ROC=0.595). O facto do modelo que apresentou melhores resultados, tendo em conta as métricas estudadas, ter acontecido no cenário S8, em que estão todas as variáveis presentes menos a navegação, pode indicar que este opcional é menos importante para prever esta Gama e que por sua vez os consumidores desta Gama estão menos interessados nele, raciocínio que é não é apoiado pela tabela 14, mostra os resultados simplificados das escolhas dos consumidores no questionário, que indica que opcional Navegação é muito desejado pelos inquiridos que têm preferência por esta gama.

Na Gama Familiar, o algoritmo que apresentou melhores resultados foi o J48 no cenário S4 (todas as variáveis menos a cor exterior), com 58.2% de acuidade. A proporção de Vendas positivas identificadas pelo teste com uma percentagem de 32.1% (sensibilidade) e de vendas negativas de 79.8% (especificidade). E capacidade preditiva do modelo classificar as vendas positivas e as negativas é 56.9% (ROC=0.569). O facto do modelo que apresentou melhores resultados, tendo em conta as métricas estudadas, ter acontecido no cenário S4, em que estão todas as variáveis presentes menos a Cor Exterior, pode indicar que este opcional é menos importante para prever esta Gama e que por sua vez os consumidores desta Gama estão menos interessados nele.

Na Gama Luxo/Supercarro, o algoritmo que apresentou melhores resultados foi o Logistic no cenário S5 (todas as variáveis menos a cor interior), com 68.6% de acuidade. A proporção de Vendas positivas identificadas pelo teste com uma percentagem de 81.4% (sensibilidade) e de vendas negativas de 48.1% (especificidade). E capacidade preditiva do modelo classificar as vendas positivas e as negativas é 56.8% (ROC=0.568). O facto do modelo que apresentou melhores resultados, tendo em conta as métricas estudadas, ter acontecido no cenário S5, em que estão todas as variáveis presentes menos a Cor Interior, pode indicar que este opcional é menos importante para prever esta Gama e que por sua vez os consumidores desta Gama estão menos interessados nele.

Na Gama SUV/Jipe, o algoritmo que apresentou melhores resultados foi o J48 no cenário S6 (todas as variáveis menos as jantes), com 55.8% de acuidade. A proporção de Vendas positivas identificadas pelo teste com uma percentagem de 53.6% (sensibilidade) e de vendas negativas de 63% (especificidade). E capacidade preditiva do modelo classificar as vendas positivas e as negativas é 56.7 (ROC=0.567). O facto do modelo que apresentou melhores resultados, tendo em

conta as métricas estudadas, ter acontecido no cenário S6, em que estão todas as variáveis presentes menos as Jantes, pode indicar que este opcional é menos importante para prever esta Gama e que por sua vez os consumidores desta Gama estão menos interessados nele, raciocínio que pode ser apoiado pela tabela 14, mostra os resultados simplificados das escolhas dos consumidores no questionário, que também indica que o opcional Jantes não é muito desejado pelos inquiridos que têm preferência por esta gama.

5.2.4 Sugestões para cada Gama

Tendo em conta os resultados da análise manual e da análise através de técnicas de mineração, foi possível determinar algumas sugestões para melhorar o desempenho em termos de dias médios que um veículo fica em stock, para cada gama.

Na gama Citadino, através da análise manual, não foi possível adicionar algo que ajudasse a melhorar os tempos de venda, por sua vez, a análise através de técnicas de mineração sugeriu-nos que o opcional jantes, é pouco importante para os consumidores que compram esta gama, isto significa que é indiferente se os veículos desta Gama estão equipados com jantes superiores às de série.

Na gama Desportivo/Cabrio, através da análise manual, não foi possível adicionar algo que ajudasse a melhorar os tempos de venda, por sua vez, a análise através de técnicas de mineração sugeriu-nos que o opcional Navegação, é pouco importante para os consumidores que compram esta gama, isto pode significar que é indiferente se os veículos desta Gama estão equipados com Navegação ou não.

Na gama Familiar, através da análise manual, entendeu-se que a cor exterior preto e a cor interior preto, assim como os opcionais, Navegação e Sensores de Estacionamento, ajudaram a melhorar os tempos de venda, por sua vez, a análise através de técnicas de mineração sugeriu-nos que o opcional Cor Exterior, é pouco importante para os consumidores que compram esta gama, isto pode significar que é relativamente indiferente a cor exterior dos veículos nesta Gama.

Na gama SUV/Jipe, através da análise manual, entendeu-se que a cor exterior preto e a cor interior preto, assim como os opcionais, Navegação e Sensores de Estacionamento, ajudaram a melhorar os tempos de venda, por sua vez, a análise através de técnicas de mineração sugeriu-nos que o opcional Jantes, é pouco importante para os consumidores que compram esta gama,

isto significa que é indiferente se os veículos desta Gama estão equipados com jantes superiores às de série.

Na gama Luxo/Supercarro, através da análise manual, não foi possível adicionar algo que ajudasse a melhorar os tempos de venda, por sua vez, a análise através de técnicas de mineração sugeriu-nos que o opcional Cor Interior, é pouco importante para os consumidores que compram esta gama, isto pode significar que é relativamente indiferente a cor interior dos veículos nesta Gama.

6. Conclusões e Limitações

6.1 Conclusões e Limitações

Com o objetivo de ajudar todos os envolvidos nesta área de negócio, através da análise manual e da criação de modelos de previsão, tentou-se descortinar algumas regras que conseguissem ser benéficas para o negócio. A análise manual revelou que para as gamas Familiar e SUV/Jipe (representam 51% da totalidade dos dados) as preferências reveladas pelos consumidores, quando comparadas com as vendas do concessionário, apresentavam melhores resultados (tempos médios de venda mais baixos), logo podemos dizer que os opcionais aqui apontados trazem melhores resultados para ambas as gamas. A utilização de técnicas de mineração, permitiu-nos encontrar modelos que conseguissem prever corretamente vendas positivas ou negativas com uma acuidade máxima de 68.6% para a Gama Luxo/Supercarro e mínimo de 55.8 para a Gama SUV/Jipe, que dependendo da área de estudo, pode ser considerado satisfatório ou não. A mineração traz-nos ainda o resultado natural do que seria uma venda positiva ou negativa através do entendimento de quais targets é que funcionaram, como T1 foi o que obteve melhores resultados no geral, podemos dizer que uma venda positiva é feita em menos de 30 dias e uma venda negativa em mais de 30 dias.

Olhando para os números mínimos e máximos dos dias de stock, verificamos que há situações que precisam de ser validadas pelo concessionário em termos de qualidade, como o facto de existirem registos em que os dias de stock estão a zero e outros em que os dias de stock são muito elevados, como é demonstrado na tabela 16, tendo isto em conta, entende-se que este tipo de registos também pode representar uma limitação. A obtenção de dados vindos das empresas (concessionários), foi difícil, o desejo de obter dados ainda melhores, vindos dos importadores tornou-se impossível, considerou-se esta uma das maiores limitações deste projeto, entende-se que estes dados podem guardar “segredos” e vantagens competitivas, como tal percebe-se o seu difícil acesso. Uma outra restrição foi o facto de os dados recolhidos serem pouco completos, apesar do número de registos ser significativos, o número de variáveis a estudar foi relativamente reduzido.

Outra limitação deste estudo foi o número reduzido de inquiridos, apenas 100 e o facto da maioria ser da faixa etária 18-24 anos, podendo assim enviesar os resultados do estudo. Uma outra limitação é o facto de os dados recolhidos das preferências dos consumidores terem sido

recolhidas múltiplas opções de cores exteriores e interiores (até 3) e neste estudo ter sido usado apenas a primeira opção de cada inquirido.

Uma outra limitação encontrada foi o fato de não ter sido recolhida a localização geográfica dos inquiridos durante o questionário e a mesma também não ter sido fornecida a quando da recolha dos dados do concessionário, o fato de esta variável não ter sido estudada pode modificar ou enviesar os resultados finais. Assim como a sazonalidade e as épocas promocionais não terem sido fornecidas a quando da recolha dos dados do concessionário, também pode esconder algumas razões do sucesso e insucesso de venda de determinadas combinações. Outra limitação relativa aos dados do concessionário é o fato de ter sido pedida discricção, relativa às marcas em si, não permitindo assim a demonstração da total transformação dos dados fornecidos.

A última limitação a qual se teve de ceder, foi o facto de se tratar de um projeto de mestrado com as restrições de tempo que lhe são associadas, como tal não foi possível alargar a pesquisa, quer em termos teóricos, quer em termos práticos.

6.2 Trabalho Futuro

Como trabalho futuro, recomenda-se a recolha de mais dados das preferências dos consumidores e analisar esses dados com os resultados das recolhas estatísticas das preferências dos consumidores disponibilizado pela ACAP (Associação do Comércio Automóvel de Portugal), de forma a perceber se ambas as recolhas vão de encontro às preferências reais dos consumidores em Portugal. Dessa forma, possibilita perceber o facto das combinações de características escolhidas pelos inquiridos em algumas gamas não serem as que se vendem mais rapidamente. Pretende-se ainda refinar as técnicas de *Data mining* já utilizadas para extrair com mais precisão as combinações de características dos veículos que demoram menos dias a serem vendidas e também para procurar mais aspetos diferenciadores dentro de cada gama.

Bibliografia

- Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, *16*(1), 125–127.
- Arnott, D., & Pervan, G. (2012). Design science in decision support systems research: An assessment using the Hevner, March, Park, and Ram Guidelines. *Journal of the Association for Information Systems*, *13*(11), 923.
- Baltas, G., & Saridakis, C. (2013). An empirical investigation of the impact of behavioural and psychographic consumer characteristics on car preferences: An integrated model of car type choice. *Transportation Research Part A: Policy and Practice*, *54*, 92–110.
<https://doi.org/10.1016/j.tra.2013.07.007>
- Beresford, B., & Sloper, P. (2008). *Understanding the dynamics of decision-making and choice: A scoping study of key psychological theories to inform the design and analysis of the Panel Study*. Social Policy Research Unit, University of York York.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS inc*, *16*.
- Choo, S., & Mokhtarian, P. L. (2004). What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A: Policy and Practice*, *38*(3), 201–222.
- Dimitoglou, G., Adams, J. A., & Jim, C. M. (2012). Comparison of the C4. 5 and a Naïve Bayes classifier for the prediction of lung cancer survivability. *arXiv preprint arXiv:1206.1121*.
- Dunham, M. H. (2006). *Data mining: Introductory and advanced topics*. Pearson Education India.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861–874.
- Freitas, A. A. (2003). A survey of evolutionary algorithms for data mining and knowledge discovery. In *Advances in evolutionary computing* (pp. 819–845). Springer.
- Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American statistical Association*, *70*(350), 320–328.
- Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: Data mining, inference and prediction. *The Mathematical Intelligencer*, *27*(2), 83–85.
- He, H., & Garcia, E. A. (2008). Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, (9), 1263–1284.
- Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 1.
- Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications—a holistic extension to the CRISP-DM model. *Procedia CIRP*, *79*, 403–408.

- Jiawei Han, M. K., & Pei, J. (2011). *Data mining: Concepts and techniques: concepts and techniques*. Elsevier.
- Kabakchieva, D. (2012). Student performance prediction by using data mining classification algorithms. *International journal of computer science and management research*, 1(4), 686–690.
- Kantardzic, M. (2011). *Data mining: Concepts, models, methods, and algorithms*. John Wiley & Sons.
- Kitamura, R., Golob, T. F., Yamamoto, T., & Wu, G. (1999). *Accessibility and auto use in a motorized metropolis*.
- Lave, C. A., & Train, K. (1979). A disaggregate model of auto-type choice. *Transportation research part A: general*, 13(1), 1–9.
- Marques, R. F. P., Correia, A. G., & Cortez, P. (2008). *Data mining no guide terrassements routiers*.
- Mason-Jones, R., Naylor, B., & Towill, D. R. (2000). Lean, agile or leagile? Matching your supply chain to the marketplace. *International Journal of Production Research*, 38(17), 4061–4070.
- Moro, S., Laureano, R., & Cortez, P. (2011). Using data mining for bank direct marketing: An application of the crisp-dm methodology. *Proceedings of European Simulation and Modelling Conference-ESM'2011*, 117–121. EUROSIS-ETI.
- Patil, T. R., & Sherekar, S. S. (2013). Performance analysis of Naive Bayes and J48 classification algorithm for data classification. *International journal of computer science and applications*, 6(2), 256–261.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-122240302>
- Provost, F., Fawcett, T., & Kohavi, R. (1998). The Case against Accuracy Estimation while Comparing Induction Algorithms. *ICML Conference*.
- Provost, F. J., & Fawcett, T. (1997). Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. *KDD*, 97, 43–48.
- Provost, Foster, & Kohavi, R. (1998). Glossary of terms. *Journal of Machine Learning*, 30(2–3), 271–274.
- Sahoo, G., & Kumar, Y. (2012). Analysis of parametric & non parametric classifiers for classification technique using WEKA. *International Journal of Information Technology and Computer Science (IJITCS)*, 4(7), 43.
- Satake, I., Xin, J. H., Tianming, T., Hansuebsai, A., Ando, K., Sato, T., ... Ohsawa, S. (2011). A comparative study of the emotional assessment of automotive exterior colors in Asia. *Progress in Organic Coatings*, 72(3), 528–540.

- Sharma, A. K., & Sahni, S. (2011). A comparative study of classification algorithms for spam email data analysis. *International Journal on Computer Science and Engineering*, 3(5), 1890–1895.
- Sliburyte, L., & Skeryte, I. (2014). What we know about consumers' color perception. *Procedia-Social and Behavioral Sciences*, 156, 468–472.
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2), 111–133.
- Vrkljan, B. H., & Anaby, D. (2011). What vehicle features are considered important when buying an automobile? An examination of driver preferences by age and gender. *Journal of safety research*, 42(1), 61–65.
- Wan, C. (2018). *Hierarchical Feature Selection for Knowledge Discovery: Application of Data Mining to the Biology of Ageing*. Springer.
- Wilkinson, S. (1998). Focus group methodology: A review. *International journal of social research methodology*, 1(3), 181–203.
- Yadav, S., & Shukla, S. (2016). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 78–83. IEEE.
- Zhang, Y., & Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, 187(1), 95–112.

Anexo I

Tabela 24 – Resultados da análise aos dados do concessionário na gama cidadão para o Target T2

T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S1	TP Rate	0	0	0	0.997	0.515	0	0	0	1	0.516	0	0	0	0.996	0.514
	Precision	?	0	0	0.516	?	?	?	?	0.516	?	?	0	0	0.515	?
	ROC	0.560	0.543	0.561	0.579	0.566	0.500	0.496	0.498	0.498	0.498	0.557	0.544	0.559	0.544	0.566
	Acuidade	0.515					0.516					0.514				
S2	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
	Precision	?	?	?	0.516	0.516	?	?	?	?	0.516	?	?	?	?	0.516
	ROC	0.483	0.520	0.554	0.552	0.540	0.500	0.496	0.498	0.498	0.498	0.476	0.520	0.550	0.547	0.536
	Acuidade	0.516					0.516					0.516				
S3	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
	Precision	?	?	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.565	0.515	0.522	0.551	0.539	0.500	0.496	0.498	0.498	0.498	0.566	0.522	0.519	0.549	0.539
	Acuidade	0.516					0.516					0.516				
S4	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
	Precision	?	?	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.552	0.534	0.554	0.559	0.552	0.500	0.496	0.498	0.498	0.498	0.556	0.526	0.552	0.564	0.553
	Acuidade	0.516					0.516					0.516				
T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S5	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
	Precision	?	?	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.569	0.544	0.540	0.565	0.556	0.500	0.496	0.498	0.498	0.498	0.570	0.548	0.541	0.564	0.557
	Acuidade	0.516					0.516					0.516				
S6	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
	Precision	?	0	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.483	0.505	0.562	0.556	0.541	0.500	0.496	0.498	0.498	0.498	0.474	0.506	0.560	0.553	0.538
	Acuidade	0.516					0.516					0.516				

S7	TP Rate	0	0	0	0.997	0.514	0	0	0	1	0.516	0	0	0	0.998	0.515
	Precision	?	0	?	0.516	?	?	?	?	0.516	?	?	0	?	0.516	?
	ROC	0.556	0.547	0.557	0.577	0.565	0.500	0.496	0.498	0.498	0.498	0.553	0.547	0.553	0.576	0.563
	Acuidade	0.514					0.516					0.516				
S8	TP Rate	0	0	0	0.997	0.514	0	0	0	1	0.516	0	0	0	0.995	0.513
	Precision	?	0	?	0.516	?	?	?	?	0.516	?	?	?	?	0.515	?
	ROC	0.562	0.545	0.558	0.581	0.567	0.500	0.496	0.498	0.498	0.498	0.562	0.543	0.556	0.581	0.567
	Acuidade	0.514					0.516					0.513				
S9	TP Rate	0	0	0	0.997	0.514	0	0	0	1	0.516	0	0	0.002	0.996	0.514
	Precision	?	0	0	0.516	?	?	?	?	0.516	?	?	?	?	0.515	?
	ROC	0.561	0.550	0.563	0.581	0.569	0.500	0.496	0.498	0.498	0.498	0.557	0.550	0.559	0.579	0.567
	Acuidade	0.514					0.516					0.514				
T2	Naives Bayes					J48					Logistic					
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
	Precision	?	?	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.502	0.508	0.528	0.508	0.512	0.500	0.496	0.498	0.498	0.498	0.499	0.508	0.528	0.508	0.511
	Acuidade	0.516					0.516					0.516				
S11	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
	Precision	?	?	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.428	0.496	0.539	0.528	0.516	0.500	0.496	0.498	0.498	0.498	0.424	0.496	0.538	0.528	0.515
	Acuidade	0.516					0.516					0.516				
S12	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
	Precision	?	?	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.557	0.534	0.489	0.535	0.528	0.500	0.496	0.498	0.498	0.498	0.557	0.534	0.489	0.535	0.528
	Acuidade	0.516					0.516					0.516				
S13	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
	Precision	?	?	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.490	0.468	0.500	0.502	0.493	0.500	0.496	0.498	0.498	0.498	0.490	0.468	0.500	0.502	0.493
	Acuidade	0.516					0.516					0.516				
S14	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516

	Precis ion	?	?	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.471	0.489	0.501	0.494	0.494	0.500	0.496	0.498	0.498	0.498	0.477	0.489	0.501	0.494	0.493
	Acuid ade	0.516					0.516					0.516				
	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516
S15	Precis ion	?	?	?	0.516	?	?	?	?	0.516	?	?	?	?	0.516	?
	ROC	0.490	0.439	0.489	0.477	0.472	0.500	0.496	0.498	0.498	0.498	0.490	0.439	0.489	0.479	0.473
	Acuid ade	0.516					0.516					0.516				
	TP Rate	0	0	0	1	0.516	0	0	0	1	0.516	0	0	0	1	0.516

Tabela 25 – Resultados da análise aos dados do concessionário na gama citadino para o Target T3

T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S1	TP Rate	0.011	0.426	0	0.673	0.375	0.050	0.391	0.013	0.660	0.369	0.011	0.442	0	0.675	0.381
	Precis ion	0.133	0.330	?	0.411	?	0.224	0.322	0.148	0.414	0.311	0.138	0.344	?	0.411	?
	ROC	0.580	0.552	0.530	0.609	0.575	0.561	0.521	0.509	0.589	0.551	0.581	0.556	0.531	0.610	0.576
	Acuid ade	0.375					0.369					0.381				
S2	TP Rate	0	0.412	0	0.691	0.376	0.005	0.420	0	0.685	0.377	0.003	0.436	0	0.666	0.374
	Precis ion	0	0.352	?	0.390	?	0.667	0.342	?	0.396	?	0.167	0.348	?	0.391	?
	ROC	0.514	0.548	0.515	0.561	0.541	0.521	0.539	0.511	0.554	0.537	0.505	0.547	0.515	0.557	0.538
	Acuid ade	0.376					0.377					0.374				
S3	TP Rate	0	0.383	0	0.740	0.385	0	0.463	0	0.677	0.386	0	0.377	0	0.748	0.386
	Precis ion	?	0.331	?	0.414	?	?	0.327	?	0.430	?	?	0.331	?	0.415	?
	ROC	0.565	0.508	0.531	0.583	0.549	0.561	0.514	0.524	0.575	0.546	0.561	0.505	0.530	0.577	0.545
	Acuid ade	0.385					0.386					0.386				
S4	TP Rate	0.005	0.396	0	0.677	0.367	0.029	0.412	0.006	0.710	0.389	0	0.391	0	0.717	0.379
	Precis ion	0.125	0.311	?	0.405	?	0.250	0.333	0.002	0.541	0.363	0	0.318	?	0.416	?
	ROC	0.569	0.534	0.523	0.595	0.561	0.581	0.529	0.524	0.607	0.566	0.578	0.532	0.520	0.597	0.562
	Acuid ade	0.367					0.389					0.379				
T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															

S5	TP Rate	0.005	0.418	0	0.699	0.382	0.011	0.334	0.044	0.726	0.373	0.008	0.399	0	0.710	0.380
	Precision	0.143	0.326	?	0.420	?	0.133	0.314	0.215	0.417	0.304	0.188	0.325	?	0.416	?
	ROC	0.579	0.540	0.536	0.601	0.569	0.559	0.494	0.536	0.582	0.544	0.577	0.539	0.538	0.600	0.568
	Acuidade	0.382					0.373					0.380				
S6	TP Rate	0.003	0.445	0	0.686	0.385	0.021	0.463	0.019	0.616	0.371	0.003	0.439	0	0.673	0.378
	Precision	0.167	0.359	?	0.401	?	0.145	0.336	0.316	0.409	0.523	0.167	0.348	?	0.397	?
	ROC	0.513	0.554	0.510	0.575	0.547	0.482	0.512	0.518	0.554	0.523	0.508	0.556	0.513	0.573	0.547
	Acuidade	0.385					0.371					0.378				
S7	TP Rate	0.003	0.410	0	0.689	0.375	0.027	0.486	0	0.597	0.369	0.005	0.393	0	0.694	0.372
	Precision	0.050	0.332	?	0.406	?	0.156	0.327	0	0.419	0.279	0.105	0.329	?	0.400	?
	ROC	0.583	0.543	0.535	0.604	0.571	0.570	0.529	0.520	0.585	0.555	0.584	0.545	0.535	0.607	0.573
	Acuidade	0.375					0.369					0.372				
S8	TP Rate	0.011	0.405	0	0.702	0.380	0.045	0.415	0.009	0.669	0.378	0.005	0.413	0	0.683	0.374
	Precision	0.133	0.332	?	0.414	?	0.029	0.350	0.010	0.442	0.378	0.077	0.334	?	0.405	?
	ROC	0.581	0.554	0.531	0.609	0.575	0.561	0.544	0.511	0.597	0.561	0.582	0.557	0.531	0.610	0.577
	Acuidade	0.380					0.378					0.374				
S9	TP Rate	0.005	0.453	0	0.670	0.382	0.029	0.418	0	0.678	0.378	0.003	0.442	0	0.674	0.379
	Precision	0.111	0.339	?	0.415	?	0.175	0.328	0	0.423	0.285	0.050	0.339	?	0.411	?
	ROC	0.579	0.556	0.534	0.612	0.577	0.553	0.525	0.521	0.602	0.557	0.580	0.559	0.534	0.610	0.578
	Acuidade	0.382					0.378					0.379				
T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0	0.237	0	0.825	0.372	0	0.164	0	0.865	0.364	0	0.251	0	0.804	0.368
	Precision	?	0.340	?	0.380	?	?	0.326	?	0.371	?	?	0.331	?	0.379	?
	ROC	0.506	0.525	0.506	0.535	0.522	0.514	0.518	0.514	0.525	0.519	0.507	0.525	0.506	0.535	0.522
	Acuidade	0.372					0.364					0.368				
S11	TP Rate	0	0.367	0	0.657	0.350	0	0.051	0	0.936	0.356	0	0.429	0	0.605	0.350
	Precision	?	0.311	?	0.372	?	?	0.283	?	0.360	?	?	0.409	?	0.376	?
	ROC	0.490	0.522	0.495	0.523	0.512	0.488	0.495	0.485	0.484	0.488	0.490	0.521	0.496	0.523	0.512
	Acuidade	0.350					0.356					0.350				

S12	TP Rate	0	0.108	0	0.882	0.353	0	0	0	1	0.363	0	0.108	0	0.882	0.353
	Precision	0	0.130	0	0.366	?	?	?	?	0.363	?	?	0.266	0	0.366	?
	ROC	0.558	0.461	0.516	0.517	0.519	0.496	0.499	0.496	0.497	0.497	0.558	0.460	0.516	0.515	0.519
	Acuidade	0.353				0.366				0.353						
S13	TP Rate	0	0	0	1	0.363	0	0	0	1	0.363	0	0	0	1	0.363
	Precision	?	?	?	0.363	?	?	?	?	0.363	?	?	?	?	0.363	?
	ROC	0.487	0.509	0.487	0.512	0.503	0.496	0.499	0.496	0.497	0.497	0.487	0.509	0.487	0.512	0.503
	Acuidade	0.363				0.363				0.363						
S14	TP Rate	0	0	0	1	0.363	0	0	0	1	0.363	0	0	0	1	0.363
	Precision	?	?	?	0.363	?	?	?	?	0.363	0.363	?	?	?	0.363	?
	ROC	0.482	0.490	0.494	0.506	0.495	0.496	0.499	0.496	0.497	0.497	0.482	0.490	0.494	0.506	0.495
	Acuidade	0.363				0.363				0.363						
S15	TP Rate	0	0	0	1	0.363	0	0	0	1	0.363	0	0	0	1	0.363
	Precision	?	?	?	0.363	?	?	?	?	0.363	0.363	?	?	?	0.363	?
	ROC	0.497	0.488	0.457	0.478	0.481	0.496	0.499	0.496	0.497	0.497	0.497	0.488	0.457	0.478	0.481
	Acuidade	0.363				0.363				0.363						

Tabela 26 – Resultados da análise aos dados do concessionário na gama cidadão para o Target T4

T4	Cenários	Classes	Naives Bayes				J48				Logistic			
			A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
S1	Métricas	TP Rate	0.992	0	0.030	0.638	0.993	0.003	0.009	0.635	0.992	0	0.032	0.639
		Precision	0.640	?	0.464	?	0.638	0.167	0.364	0.508	0.641	?	0.452	?
		ROC	0.615	0.558	0.598	0.603	0.517	0.502	0.524	0.517	0.617	0.562	0.96	0.604
		Acuidade	0.683				0.635				0.639			
S2	Métricas	TP Rate	1	0	0	0.637	1	0	0	0.637	1	0	0	0.637
		Precision	0.637	?	?	?	0.637	?	?	?	0.637	?	?	?
		ROC	0.569	0.553	0.529	0.558	0.498	0.497	0.497	0.497	0.568	0.552	0.523	0.556
		Acuidade	0.637				0.637				0.637			
S3	Métricas	TP Rate	1	0	0	0.637	1	0	0	0.637	1	0	0	0.637
		Precision	0.637	?	?	?	0.637	?	?	?	0.637	?	?	?

	ROC	0.584	0.516	0.589	0.574	0.498	0.497	0.497	0.497	0.580	0.520	0.586	0.572
	Acuidad e	0.637				0.637				0.637			
S4	TP Rate	0.995	0	0.016	0.637	1	0	0	0.637	0.995	0	0.021	0.638
	Precisio n	0.639	?	0.412	?	0.637	?	?	?	0.639	?	0.500	?
	ROC	0.600	0.539	0.595	0.590	0.498	0.497	0.497	0.497	0.602	0.545	0.595	0.592
	Acuidad e	0.637				0.637				0.638			
T4		Naives Bayes				J48				Logistic			
Cenários	Classes	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
S5	TP Rate	0.992	0	0.007	0.633	1	0	0	0.637	0.994	0	0.002	0.633
	Precisio n	0.637	?	0.176	?	0.637	?	?	?	0.636	?	0.111	?
	ROC	0.603	0.543	0.593	0.592	0.498	0.497	0.497	0.497	0.600	0.546	0.589	0.590
	Acuidad e	0.633				0.637				0.633			
S6	TP Rate	0.998	0	0.002	0.636	1	0	0	0.637	0.998	0	0.005	0.637
	Precisio n	0.637	?	0.200	?	0.637	?	?	?	0.637	?	0.002	?
	ROC	0.580	0.546	0.955	0.570	0.498	0.497	0.497	0.497	0.579	0.542	0.552	0.568
	Acuidad e	0.636				0.637				0.637			
S7	TP Rate	0.994	0	0.030	0.639	0.994	0.003	0.007	0.635	0.996	0	0.030	0.641
	Precisio n	0.640	?	0.542	?	0.637	0.167	0.333	0.501	0.641	?	0.619	?
	ROC	0.611	0.560	0.593	0.600	0.517	0.502	0.526	0.517	0.614	0.561	0.591	0.601
	Acuidad e	0.639				0.635				0.641			
S8	TP Rate	0.996	0	0.028	0.640	1	0	0	0.637	0.994	0	0.037	0.641
	Precisio n	0.640	?	0.600	?	0.637	?	?	?	0.642	?	0.516	?
	ROC	0.617	0.562	0.598	0.604	0.498	0.497	0.497	0.497	0.619	0.183	0.597	0.606
	Acuidad e	0.640				0.637				0.641			
S9	TP Rate	0.992	0	0.032	0.639	1	0	0	0.637	0.992	0	0.035	0.639
	Precisio n	0.640	?	0.500	?	0.637	?	?	?	0.641	?	0.484	?
	ROC	0.617	0.562	0.598	0.605	0.498	0.497	0.497	0.497	0.618	0.565	0.597	0.605
	Acuidad e	0.639				0.637				0.639			
T4		Naives Bayes				J48				Logistic			
U	Classes	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg

	Métricas												
S10	TP Rate	1	0	0	0.637	1	0	0	0.637	1	0	0	0.637
	Precisión	0.637	?	?	?	0.637	?	?	?	0.637	?	?	?
	ROC	0.539	0.519	0.479	0.524	0.498	0.497	0.497	0.497	0.540	0.519	0.479	0.524
	Acuidad e	0.637				0.637				0.637			
S11	TP Rate	1	0	0	0.637	1	0	0	0.637	1	0	0	0.637
	Precisión	0.637	?	?	?	0.637	?	?	?	0.637	?	?	?
	ROC	0.539	0.527	0.528	0.535	0.498	0.497	0.497	0.497	0.537	0.526	0.524	0.533
	Acuidad e	0.637				0.637				0.637			
S12	TP Rate	1	0	0	0.637	1	0	0	0.637	1	0	0	0.637
	Precisión	0.637	?	?	0.637	0.637	?	?	?	0.637	?	?	?
	ROC	0.561	0.508	0.561	0.553	0.498	0.497	0.497	0.497	0.561	0.508	0.561	0.553
	Acuidad e	0.637				0.637				0.637			
S13	TP Rate	1	0	0	0.637	1	0	0	0.637	1	0	0	0.637
	Precisión	0.637	?	?	?	0.637	?	?	?	0.637	?	?	?
	ROC	0.507	0.488	0.502	0.503	0.498	0.497	0.497	0.497	0.507	0.488	0.502	0.503
	Acuidad e	0.637				0.637				0.637			
S14	TP Rate	1	0	0	0.637	1	0	0	0.637	1	0	0	0.637
	Precisión	0.637	?	?	?	0.637	?	?	?	0.637	?	?	?
	ROC	0.502	0.492	0.495	0.499	0.498	0.497	0.497	0.497	0.502	0.492	0.495	0.499
	Acuidad e	0.637				0.637				0.637			
S15	TP Rate	1	0	0	0.637	1	0	0	0.637	1	0	0	0.637
	Precisión	0.637	?	?	?	0.637	?	?	?	0.637	?	?	?
	ROC	0.482	0.446	0.492	0.479	0.498	0.497	0.497	0.497	0.482	0.448	0.492	0.479
	Acuidad e	0.637				0.637				0.637			

Anexo II

Tabela 27 – Resultados da análise aos dados do concessionário na gama Desportivo/Cabrio para o Target T2

T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S1	TP Rate	0	0.374	0	0.732	0.411	0.032	0.198	0	0.691	0.344	0	0.385	0	0.707	0.404
	Precision	0	0.386	0	0.448	0.299	0.333	0.228	0	0.407	0.269	0	0.365	0	0.455	0.295
	ROC	0.464	0.560	0.326	0.562	0.507	0.445	0.452	0.336	0.453	0.430	0.485	0.553	0.277	0.570	0.501
	Acuidade	0.411					0.344					0.404				
S2	TP Rate	0	0.242	0	0.740	0.374	0	0.143	0	0.829	0.381	0	0.396	0	0.675	0.394
	Precision	?	0.275	0	0.412	?	?	0.255	?	0.406	?	?	0.336	0	0.435	?
	ROC	0.414	0.560	0.295	0.526	0.481	0.456	0.470	0.339	0.470	0.444	0.459	0.539	0.270	0.538	0.479
	Acuidade	0.374					0.381					0.394				
S3	TP Rate	0	0.165	0	0.821	0.384	0	0.066	0	0.870	0.488	0	0.176	0	0.854	0.401
	Precision	?	0.278	?	0.407	?	?	0.273	0.100	0.396	?	?	0.327	?	0.415	?
	ROC	0.500	0.464	0.461	0.538	0.497	0.450	0.481	0.477	0.499	0.484	0.502	0.476	0.445	0.531	0.495
	Acuidade	0.384					0.488					0.401				
S4	TP Rate	0	0.286	0	0.797	0.411	0	0.198	0	0.829	0.397	0	0.253	0	0.821	0.411
	Precision	?	0.366	?	0.424	?	?	0.340	0	0.415	?	?	0.343	?	0.430	?
	ROC	0.428	0.498	0.368	0.546	0.486	0.362	0.453	0.406	0.478	0.445	0.452	0.501	0.338	0.547	0.484
	Acuidade	0.411					0.397					0.411				
T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S5	TP Rate	0	0.330	0	0.780	0.417	0	0.231	0.018	0.715	0.364	0	0.385	0	0.764	0.427
	Precision	0	0.380	0	0.438	0.293	0	0.323	0.043	0.415	0.275	0	0.407	0	0.441	0.302
	ROC	0.514	0.539	0.397	0.548	0.513	0.438	0.514	0.417	0.521	0.491	0.514	0.538	0.367	0.547	0.507
	Acuidade	0.417					0.364					0.427				
S6	TP Rate	0	0.396	0	0.732	0.417	0.065	0.198	0	0.740	0.368	0	0.396	0	0.724	0.414
	Precision	0	0.383	0	0.455	0.301	0.667	0.243	0	0.412	0.368	0	0.371	0	0.464	0.301
	ROC	0.473	0.563	0.338	0.573	0.516	0.453	0.460	0.301	0.461	0.430	0.466	0.556	0.282	0.579	0.504

	Acuidade	0.417					0.368					0.414				
S7	TP Rate	0	0.330	0	0.675	0.374	0.032	0.231	0	0.675	0.348	0	0.396	0	0.626	0.474
	Precision	0	0.333	0	0.403	0.265	0.250	0.273	0	0.395	0.269	0	0.333	0	0.418	0.271
	ROC	0.472	0.564	0.329	0.536	0.499	0.432	0.490	0.332	0.447	0.437	0.497	0.544	0.293	0.553	0.496
	Acuidade	0.374					0.348					0.374				
S8	TP Rate	0	0.352	0	0.764	0.417	0	0.209	0	0.740	0.364	0	0.420	0	0.691	0.411
	Precision	?	0.352	0	0.452	?	0	0.241	0	0.419	0.243	0	0.368	0	0.457	0.297
	ROC	0.455	0.555	0.316	0.567	0.504	0.394	0.428	0.270	0.448	0.403	0.484	0.553	0.283	0.574	0.503
	Acuidade	0.417					0.364					0.411				
S9	TP Rate	0	0.385	0	0.780	0.434	0.032	0.209	0	0.699	0.351	0	0.374	0	0.732	0.411
	Precision	?	0.402	0	0.459	?	0.333	0.238	0	0.402	0.269	0	0.366	0	0.455	0.295
	ROC	0.443	0.567	0.305	0.552	0.499	0.414	0.454	0.306	0.439	0.416	0.487	0.561	0.262	0.563	0.498
	Acuidade	0.434					0.351					0.411				
T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0	0.462	0	0.732	0.437	0	0.044	0	0.951	0.401	0	0.462	0	0.732	0.437
	Precision	?	0.404	?	0.455	?	?	0.308	?	0.405	?	?	0.404	?	0.455	?
	ROC	0.435	0.543	0.389	0.497	0.484	0.510	0.496	0.467	0.476	0.484	0.441	0.543	0.388	0.498	0.485
	Acuidade	0.437					0.401					0.437				
S11 S12	TP Rate	0	0.308	0	0.813	0.422	0	0.242	0	0.829	0.411	0	0.308	0	0.813	0.422
	Precision	?	0.378	?	0.439	?	?	0.333	?	0.432	?	?	0.378	?	0.439	?
	ROC	0.375	0.472	0.266	0.491	0.431	0.348	0.469	0.287	0.489	0.430	0.345	0.471	0.272	0.481	0.425
	Acuidade	0.422					0.411					0.424				
	TP Rate	0	0	0	1	0.407	0	0	0	1	0.407	0	0	0	1	0.407
	Precision	?	?	?	0.407	?	?	?	?	0.407	?	?	?	?	0.407	?
	ROC	0.431	0.422	0.384	0.441	0.423	0.484	0.493	0.475	0.484	0.485	0.431	0.422	0.384	0.448	0.426
	Acuidade	0.407					0.407					0.407				
S13	TP Rate	0	0.033	0	0.935	0.391	0	0	0	1	0.407	0	0.033	0	0.935	0.391
	Precision	?	0.167	?	0.405	?	?	?	?	0.407	?	?	0.167	?	0.405	?
	ROC	0.452	0.463	0.444	0.513	0.479	0.484	0.493	0.475	0.484	0.485	0.452	0.463	0.444	0.513	0.479

	Acuidade	0.391					0.407					0.391				
S14	TP Rate	0	0	0	1	0.407	0	0	0	1	0.407	0	0	0	1	0.407
	Precisão	?	?	?	0.407	?	?	?	?	0.407	?	?	?	?	0.407	?
	ROC	0.483	0.497	0.468	0.421	0.459	0.484	0.493	0.475	0.484	0.485	0.483	0.497	0.468	0.420	0.458
	Acuidade	0.407					0.407					0.407				
S15	TP Rate	0	0	0	1	0.407	0	0	0	1	0.407	0	0	0	1	0.407
	Precisão	?	?	?	0.407	?	?	?	?	0.407	?	?	?	?	0.407	?
	ROC	0.512	0.446	0.491	0.501	0.484	0.484	0.493	0.475	0.484	0.485	0.512	0.454	0.491	0.501	0.486
	Acuidade	0.407					0.407					0.407				

Tabela 28 – Resultados da análise aos dados do concessionário na gama Desportivo/Cabrio para o Target T4

T4		Naives Bayes				J48				Logistic			
Cenários	Classes	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
S1	TP Rate	0.996	0	0.020	0.742	1	0	0	0.742	0.987	0	0	0.732
	Precisão	0.743	?	0.500	?	0.742	?	?	?	0.739	?	0	?
	ROC	0.516	0.646	0.545	0.532	0.484	0.455	0.489	0.483	0.497	0.668	0.542	0.520
	Acuidade	0.742				0.742				0.732			
S2	TP Rate	1	0	0	0.742	1	0	0	0.742	1	0	0	0.742
	Precisão	0.742	?	?	?	0.742	?	?	?	0.742	?	?	?
	ROC	0.506	0.563	0.487	0.508	0.484	0.455	0.489	0.483	0.494	0.575	0.433	0.491
	Acuidade	0.742				0.742				0.742			
S3	TP Rate	1	0	0	0.742	1	0	0	0.742	1	0	0	0.742
	Precisão	0.742	?	?	?	0.742	?	?	?	0.742	?	?	?
	ROC	0.506	0.610	0.580	0.528	0.484	0.455	0.489	0.483	0.500	0.620	0.584	0.525
	Acuidade	0.742				0.742				0.742			
S4	TP Rate	1	0	0	0.742	1	0	0	0.742	0.996	0	0	0.738
	Precisão	0.742	?	?	?	0.742	?	?	?	0.741	?	0	?
	ROC	0.475	0.630	0.571	0.505	0.484	0.455	0.489	0.483	0.455	0.634	0.576	0.492
	Acuidade	0.742				0.742				0.738			
T4		Naives Bayes				J48				Logistic			
U	Classes	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg

	Métricas												
S5	TP Rate	1	0	0	0.742	1	0	0	0.742	1	0	0	0.742
	Precisión	0.742	?	?	?	0.742	?	?	?	0.742	?	?	?
	ROC	0.536	0.624	0.541	0.545	0.484	0.455	0.489	0.483	0.526	0.622	0.567	0.542
	Acuidad e	0.742				0.742				0.742			
S6	TP Rate	0.996	0	0	0.738	1	0	0	0.742	0.996	0	0	0.738
	Precisión	0.741	?	0	?	0.742	?	?	?	0.741	?	0	?
	ROC	0.523	0.646	0.546	0.538	0.484	0.455	0.489	0.483	0.501	0.666	0.544	0.523
	Acuidad e	0.738				0.742				0.738			
S7	TP Rate	0.996	0	0.020	0.742	1	0	0	0.742	1	0	0	0.742
	Precisión	0.743	?	0.500	?	0.742	?	?	?	0.742	?	?	?
	ROC	0.503	0.518	0.537	0.510	0.484	0.455	0.489	0.483	0.497	0.557	0.520	0.506
	Acuidad e	0.742				0.742				0.742			
S8	TP Rate	0.996	0	0.020	0.742	1	0	0	0.742	0.991	0	0	0.735
	Precisión	0.743	?	0.500	?	0.742	?	?	?	0.740	?	0	?
	ROC	0.535	0.645	0.536	0.545	0.484	0.455	0.489	0.483	0.513	0.661	0.539	0.531
	Acuidad e	0.742				0.742				0.735			
S9	TP Rate	0.982	0	0	0.728	1	0	0	0.742	0.987	0	0	0.732
	Precisión	0.738	?	0	?	0.742	?	?	?	0.739	?	0	?
	ROC	0.500	0.662	0.524	0.518	0.484	0.455	0.489	0.483	0.480	0.681	0.491	0.500
	Acuidad e	0.728				0.742				0.732			
T4		Naives Bayes				J48				Logistic			
Cenári	Clases	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
S10	TP Rate	1	0	0	0.742	1	0	0	0.742	1	0	0	0.742
	Precisión	0.742	?	?	?	0.742	?	?	?	0.742	?	?	?
	ROC	0.520	0.502	0.455	0.508	0.484	0.455	0.489	0.483	0.515	0.512	0.449	0.503
	Acuidad e	0.742				0.742				0.742			
S11	TP Rate	1	0	0	0.742	1	0	0	0.742	1	0	0	0.742
	Precisión	0.742	?	?	?	0.742	?	?	?	0.742	?	?	?
	ROC	0.436	0.522	0.509	0.456	0.484	0.455	0.489	0.483	0.450	0.531	0.487	0.463

	Acuidad e	0.742				0.742				0.742			
S1 2	TP Rate	1	0	0	0.74 2	1	0	0	0.74 2	1	0	0	0.74 2
	Precisio n	0.74 2	?	?	?	0.74 2	?	?	?	0.74 2	?	?	?
	ROC	0.47 0	0.44 9	0.51 2	0.47 5	0.48 4	0.45 5	0.48 9	0.48 3	0.47 0	0.44 9	0.51 2	0.47 5
	Acuidad e	0.742				0.742				0.742			
S1 3	TP Rate	1	0	0	0.74 2	1	0	0	0.74 2	1	0	0	0.74 2
	Precisio n	0.74 2	?	?	?	0.74 2	?	?	?	0.74 2	?	?	?
	ROC	0.50 1	0.57 9	0.50 5	0.50 8	0.48 4	0.45 5	0.48 9	0.48 3	0.50 0	0.57 9	0.50 5	0.50 8
	Acuidad e	0.742				0.742				0.742			
S1 4	TP Rate	1	0	0	0.74 2	1	0	0	0.74 2	1	0	0	0.74 2
	Precisio n	0.74 2	?	?	?	0.74 2	?	?	?	0.74 2	?	?	?
	ROC	0.36 9	0.44 0	0.47 2	0.39 2	0.48 4	0.45 5	0.48 9	0.48 3	0.36 9	0.44 0	0.47 2	0.39 3
	Acuidad e	0.742				0.742				0.742			
S1 5	TP Rate	1	0	0	0.74 2	1	0	0	0.74 2	1	0	0	0.74 2
	Precisio n	0.74 2	?	?	?	0.74 2	?	?	?	0.74 2	?	?	?
	ROC	0.48 7	0.30 1	0.51 4	0.47 5	0.48 4	0.45 5	0.48 9	0.48 3	0.48 7	0.29 9	0.51 4	0.47 5
	Acuidad e	0.742				0.742				0.742			

Anexo III

Tabela 29 – Resultados da análise aos dados do concessionário na gama Familiar para o Target T2

T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S1	TP Rate	0	0.045	0.031	0.964	0.463	0.006	0.003	0.008	0.989	0.461	0	0.045	0.010	0.974	0.463
	Precision	?	0.315	0.250	0.474	?	1	0.333	0.176	0.464	?	?	0.340	0.167	0.471	?
	ROC	0.497	0.540	0.520	0.562	0.542	0.488	0.513	0.493	0.497	0.499	0.489	0.538	0.516	0.560	0.538
	Acuidade	0.463					0.461					0.463				
S2	TP Rate	0	0	0.010	0.990	0.461	0	0	0	1	0.463	0	0	0.008	0.990	0.460
	Precision	?	0	0.286	0.464	?	?	?	?	0.463	?	?	0	0.231	0.464	0.510
	ROC	0.460	0.519	0.511	0.520	0.512	0.492	0.496	0.496	0.497	0.496	0.451	0.519	0.508	0.519	0.510
	Acuidade	0.461					0.463					0.460				
S3	TP Rate	0	0.003	0.008	0.986	0.459	0	0	0	1	0.463	0	0	0	1	0.463
	Precision	?	0.200	0.158	0.463	?	?	?	?	0.463	?	?	?	?	0.463	?
	ROC	0.513	0.531	0.515	0.513	0.535	0.492	0.496	0.496	0.497	0.496	0.520	0.530	0.509	0.549	0.533
	Acuidade	0.459					0.463					0.463				
S4	TP Rate	0	0.016	0.013	0.989	0.464	0	0	0	1	0.463	0	0.003	0.010	0.994	0.463
	Precision	?	0.333	0.294	0.467	?	?	?	?	0.463	?	?	0.111	0.308	0.466	?
	ROC	0.504	0.527	0.507	0.558	0.534	0.492	0.496	0.496	0.497	0.496	0.504	0.526	0.505	0.551	0.531
	Acuidade	0.464					0.463					0.463				
T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S5	TP Rate	0	0.040	0.013	0.965	0.459	0	0	0	1	0.463	0	0.021	0.008	0.986	0.541
	Precision	?	0.306	0.143	0.470	?	?	?	?	0.463	?	?	0.400	0.150	0.468	?
	ROC	0.508	0.541	0.527	0.563	0.545	0.492	0.496	0.496	0.497	0.496	0.506	0.539	0.522	0.559	0.541
	Acuidade	0.459					0.463					0.541				
S6	TP Rate	0	0.019	0.013	0.976	0.459	0	0	0	1	0.463	0	0.021	0.010	0.988	0.464
	Precision	?	0.250	0.185	0.467	?	?	?	?	0.463	?	?	0.471	0.211	0.467	?
	ROC	0.496	0.539	0.523	0.557	0.540	0.492	0.496	0.496	0.497	0.496	0.488	0.538	0.521	0.550	0.535
	Acuidade	0.459					0.463					0.464				

S7	TP Rate	0	0.035	0.005	0.974	0.460	0	0	0	1	0.463	0	0.040	0.008	0.974	0.462
	Precision	?	0.271	0.105	0.469	?	?	?	?	0.463	?	?	0.283	0.150	0.471	?
	ROC	0.493	0.541	0.518	0.555	0.538	0.492	0.496	0.496	0.497	0.496	0.488	0.540	0.515	0.555	0.536
	Acuidade	0.460					0.463					0.462				
S8	TP Rate	0	0.003	0.018	0.984	0.460	0	0	0	1	0.463	0	0.005	0.013	0.984	0.460
	Precision	?	0.071	0.226	0.468	?	?	?	?	0.463	?	?	0.133	0.208	0.466	?
	ROC	0.492	0.527	0.519	0.549	0.532	0.492	0.496	0.496	0.497	0.496	0.485	0.528	0.516	0.545	0.529
	Acuidade	0.460					0.463					0.460				
S9	TP Rate	0	0.040	0.025	0.962	0.460	0.006	0.003	0.008	0.988	0.460	0	0.010	0.040	0.975	0.463
	Precision	?	0.294	0.222	0.471	?	1	0.143	0.167	0.464	0.379	?	0.349	0.133	0.472	?
	ROC	0.481	0.538	0.523	0.567	0.542	0.487	0.515	0.492	0.500	0.501	0.461	0.539	0.520	0.564	0.539
	Acuidade	0.460					0.460					0.463				
T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0	0	0.008	0.993	0.462	0	0	0	1	0.463	0	0	0.008	0.993	0.462
	Precision	?	0.214	?	0.464	?	?	?	?	0.463	?	?	0.214	0.464	?	
	ROC	0.482	0.513	0.514	0.511	0.509	0.492	0.496	0.496	0.497	0.496	0.477	0.511	0.514	0.511	0.509
	Acuidade	0.462					0.463					0.462				
S11	TP Rate	0	0	0	1	0.463	0	0	0	1	0.463	0	0	0	1	0.463
	Precision	?	?	?	0.463	?	?	?	?	0.463	?	?	?	?	0.463	?
	ROC	0.458	0.498	0.486	0.489	0.487	0.492	0.496	0.496	0.497	0.496	0.458	0.496	0.486	0.489	0.487
	Acuidade	0.463					0.463					0.463				
S12	TP Rate	0	0	0	1	0.463	0	0	0	1	0.463	0	0	0	1	0.463
	Precision	?	?	?	0.463	?	?	?	?	0.463	?	?	?	?	0.463	?
	ROC	0.483	0.497	0.491	0.521	0.505	0.492	0.496	0.496	0.497	0.496	0.483	0.497	0.491	0.521	0.505
	Acuidade	0.463					0.463					0.463				
S13	TP Rate	0	0	0	1	0.463	0	0	0	1	0.463	0	0	0	1	0.463
	Precision	?	?	?	0.463	?	?	?	?	0.463	?	?	?	?	0.463	?
	ROC	0.493	0.494	0.500	0.515	0.505	0.492	0.496	0.496	0.497	0.496	0.493	0.494	0.500	0.515	0.505
	Acuidade	0.463					0.463					0.463				

S14	TP Rate	0	0	0	1	0.463	0	0	0	1	0.463	0	0	0	1	0.463
	Precision	?	?	?	0.463	?	?	?	?	0.463	?	?	?	?	0.463	?
	ROC	0.488	0.515	0.501	0.531	0.517	0.492	0.496	0.496	0.497	0.496	0.488	0.515	0.501	0.531	0.517
	Acuidade	0.463					0.463					0.463				
S15	TP Rate	0	0	0	1	0.463	0	0	0	1	0.463	0	0	0	1	0.463
	Precision	?	?	?	0.463	?	?	?	?	0.463	?	?	?	?	0.463	?
	ROC	0.497	0.497	0.490	0.708	0.487	0.492	0.496	0.496	0.497	0.496	0.497	0.497	0.490	0.708	0.487
	Acuidade	0.463					0.463					0.463				

Tabela 30 – Resultados da análise aos dados do concessionário na gama Familiar para o Target T3

T3	Cenários	Naives Bayes					J48					Logistic				
		Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D
S1	TP Rate	0.003	0.555	0	9.551	0.363	0.064	0.529	0.016	0.540	0.366	0.008	0.573	0	0.549	0.369
	Precision	0.125	?	0.354	0.375	?	0.324	0.363	0.235	0.374	0.341	0.333	0.360	?	0.380	?
	ROC	0.491	0.560	0.537	0.595	0.554	0.512	0.544	0.529	0.589	0.550	0.493	0.560	0.535	0.593	0.553
	Acuidade	0.363					0.363					0.369				
S2	TP Rate	0	0.614	0	0.462	0.353	0.006	0.535	0	0.540	0.353	0	0.600	0	0.483	0.355
	Precision	?	0.354	?	0.352	?	0.154	0.356	?	0.353	?	?	0.357	?	0.353	?
	ROC	0.494	0.538	0.500	0.537	0.523	0.482	0.510	0.512	0.524	0.509	0.496	0.536	0.507	0.539	0.525
	Acuidade	0.353					0.353					0.355				
S3	TP Rate	0	0.557	0	0.544	0.361	0.003	0.536	0	0.560	0.356	0	0.543	0	0.549	0.358
	Precision	?	0.352	?	0.370	?	0.083	0.352	?	0.363	?	?	0.350	?	0.366	?
	ROC	0.492	0.546	0.540	0.579	0.545	0.492	0.529	0.553	0.573	0.536	0.488	0.542	0.532	0.578	0.541
	Acuidade	0.361					0.356					0.358				
S4	TP Rate	0	0.528	0	0.565	0.358	0.025	0.595	0.004	0.531	0.375	0	0.606	0	0.524	0.370
	Precision	?	0.351	?	0.364	?	0.346	0.369	0.056	0.391	0.327	?	0.362	?	0.380	?
	ROC	0.488	0.553	0.528	0.590	0.548	0.481	0.553	0.518	0.604	0.550	0.487	0.549	0.523	0.589	0.546
	Acuidade	0.358					0.375					0.370				
T3		Naives Bayes					J48					Logistic				
Ce	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg

	Métricas															
S5	TP Rate	0.003	0.574	0	0.552	0.370	0.003	0.531	0.008	0.522	0.347	0.003	0.562	0	0.549	0.365
	Precision	0.143	0.366	?	0.376	?	0.040	0.342	0.286	0.362	0.279	0.125	0.359	?	0.373	?
	ROC	0.498	0.557	0.550	0.591	0.555	0.505	0.529	0.530	0.546	0.530	0.497	0.556	0.547	0.586	0.552
	Acuidade	0.370					0.347					0.365				
S6	TP Rate	0.003	0.590	0	0.560	0.377	0.028	0.448	0	0.650	0.365	0.006	0.583	0	0.560	0.375
	Precision	0.125	0.369	?	0.369	?	0.159	0.375	0	0.372	0.277	0.222	0.501	0	0.422	?
	ROC	0.493	0.553	0.540	0.589	0.551	0.504	0.545	0.520	0.571	0.541	0.494	0.552	0.537	0.587	0.549
	Acuidade	0.377					0.365					0.375				
S7	TP Rate	0.003	0.583	0	0.535	0.367	0.020	0.493	0.008	0.512	0.334	0.003	0.580	0	0.524	0.362
	Precision	0.143	0.367	?	0.368	?	0.194	0.345	0.080	0.336	0.274	0.125	0.363	?	0.363	?
	ROC	0.488	0.561	0.536	0.586	0.551	0.495	0.518	0.502	0.535	0.516	0.495	0.560	0.534	0.585	0.551
	Acuidade	0.367					0.334					0.362				
S8	TP Rate	0	0.564	0	0.529	0.358	0.034	0.555	0.008	0.472	0.345	0.003	0.574	0	0.531	0.363
	Precision	0	0.352	?	0.367	?	0.179	0.350	0.133	0.358	0.287	0.200	0.358	?	0.370	?
	ROC	0.486	0.555	0.540	0.580	0.547	0.488	0.511	0.511	0.534	0.514	0.492	0.556	0.540	0.580	0.549
	Acuidade	0.358					0.345					0.363				
S9	TP Rate	0.003	0.571	0	0.561	0.371	0.039	0.431	0	0.670	0.367	0.006	0.583	0	0.558	0.375
	Precision	0.250	0.361	?	0.361	?	0.233	0.367	0	0.379	0.292	0.286	0.365	?	0.387	?
	ROC	0.495	0.563	0.493	0.592	0.548	0.498	0.527	0.525	0.582	0.538	0.496	0.563	0.495	0.592	0.549
	Acuidade	0.371					0.367					0.375				
T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0	0.708	0	0.384	0.359	0	0.621	0	0.444	0.350	0	0.663	0	0.416	0.354
	Precision	?	0.350	?	0.376	?	?	0.345	?	0.357	?	?	0.346	?	0.368	?
	ROC	0.486	0.521	0.513	0.532	0.516	0.488	0.521	0.512	0.532	0.517	0.488	0.521	0.512	0.532	0.517
	Acuidade	0.359					0.350					0.354				
S11	TP Rate	0	0.588	0	0.490	0.354	0	0.588	0	0.490	0.354	0	0.588	0	0.490	0.354
	Precision	?	0.352	?	0.355	?	?	0.353	?	0.355	?	?	0.337	?	0.355	?
	ROC	0.488	0.512	0.465	0.504	0.498	0.488	0.512	0.472	0.504	0.499	0.488	0.512	0.472	0.504	0.499

	Acuidade	354					354					354				
S12	TP Rate	0	0.664	0	0.451	0.366	0	0.664	0	0.451	0.366	0	0.664	0	0.451	0.366
	Precis ion	?	0.362	?	0.374	?	?	0.362	?	0.374	?	?	0.362	?	0.374	?
	ROC	0.484	0.519	0.454	0.525	0.505	0.484	0.519	0.499	0.525	0.504	0.484	0.519	0.449	0.525	0.504
	Acuidade	0.366					0.366					0.366				
S13	TP Rate	0	0.284	0	0.799	0.352	0	0.284	0	0.799	0.352	0	0.284	0	0.799	0.352
	Precis ion	?	0.377	0.344	?	?	?	0.377	?	0.344	?	?	0.377	?	0.344	?
	ROC	0.485	0.504	0.491	0.517	0.502	0.483	0.504	0.491	0.517	0.502	0.483	0.504	0.491	0.517	0.502
	Acuidade	0.352					0.352					0.352				
S14	TP Rate	0	0.370	0	0.746	0.364	0	0.370	0	0.746	0.364	0	0.370	0	0.746	0.364
	Precis ion	?	0.377	?	0.358	?	?	0.304	?	0.640	?	?	0.377	?	0.358	?
	ROC	0.487	0.517	0.481	0.532	0.511	0.487	0.517	0.481	0.532	0.511	0.487	0.517	0.481	0.532	0.511
	Acuidade	0.364					0.364					0.364				
S15	TP Rate	0	0.484	0	0.590	0.351	0	0.484	0	0.590	0.351	0	0.484	0	0.590	0.351
	Precis ion	?	0.344	?	0.358	?	?	0.344	?	0.358	?	?	0.344	?	0.358	?
	ROC	0.454	0.495	0.537	0.515	0.499	0.454	0.495	0.537	0.515	0.499	0.454	0.495	0.537	0.515	0.499
	Acuidade	0.351					0.351					0.351				

Tabela 31 – Resultados da análise aos dados do concessionário na gama Familiar para o Target T4

T4		Naives Bayes				J48				Logistic			
Cenários	Classes	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
S1	TP Rate	0.997	0	0.018	0.678	1	0	0	0.677	0.996	0	0.021	0.678
	Precis ion	0.679	?	0.545	?	0.677	?	?	?	0.679	?	0.538	?
	ROC	0.597	0.565	0.626	0.598	0.497	0.497	0.498	0.497	0.596	0.564	0.627	0.598
	Acuidade	0.678				0.677				0.678			
S2	TP Rate	1	0	0	0.667	1	0	0	0.677	1	0	0	0.677
	Precis ion	0.677	?	?	?	0.677	?	?	?	0.677	?	?	?
	ROC	0.540	0.564	0.569	0.549	0.497	0.497	0.498	0.497	0.540	0.564	0.571	0.549
	Acuidade	0.677				0.677				0.677			
S3	TP Rate	1	0	0	0.667	1	0	0	0.677	1	0	0	0.677

	Precisión	0.677	?	?	?	0.677	?	?	?	0.677	?	?	?
	ROC	0.581	0.513	0.599	0.575	0.497	0.497	0.498	0.497	0.580	0.516	0.599	0.575
	Acuidad e	0.677				0.677				0.677			
	S4	TP Rate	0.998	0	0	0.676	1	0	0	0.677	0.998	0	0
	Precisión	0.677	?	0	?	0.677	?	?	?	0.676	?	0	?
	ROC	0.587	0.531	0.622	0.586	0.497	0.497	0.498	0.497	0.588	0.526	0.623	0.586
	Acuidad e	0.676				0.677				0.676			
	T4		Naives Bayes				J48				Logistic		
Cenários	Clases	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
	TP Rate	0.985	0	0.018	0.670	1	0	0	0.677	1	0	0	0.677
	Precisión	0.678	?	0.207	?	0.677	?	?	?	0.677	?	?	?
	ROC	0.594	0.562	0.603	0.592	0.497	0.497	0.498	0.497	0.914	0.564	0.598	0.588
	Acuidad e	0.670				0.677				0.677			
	TP Rate	0.994	0	0.012	0.675	1	0	0	0.677	0.997	0	0.009	0.676
	Precisión	0.678	?	0.308	?	0.677	?	?	?	0.677	?	0.429	?
	ROC	0.589	0.564	0.617	0.591	0.497	0.497	0.498	0.497	0.587	0.564	0.618	0.590
	Acuidad e	0.675				0.677				0.676			
	TP Rate	0.998	0	0.018	0.679	1	0	0	0.677	0.998	0	0.009	0.677
	Precisión	0.679	?	0.667	?	0.677	?	?	?	0.678	?	0.500	?
	ROC	0.588	0.556	0.623	0.590	0.497	0.497	0.498	0.497	0.589	0.557	0.621	0.591
	Acuidad e	0.679				0.677				0.677			
	TP Rate	0.997	0	0.009	0.677	1	0	0	0.677	0.997	0	0.009	0.676
	Precisión	0.678	?	0.429	?	0.677	?	?	?	0.678	?	0.375	?
	ROC	0.584	0.561	0.613	0.586	0.497	0.497	0.498	0.497	0.583	0.561	0.613	0.586
	Acuidad e	0.677				0.677				0.676			
	TP Rate	0.995	0	0.006	0.675	1	0	0	0.677	0.997	0	0.006	0.676
	Precisión	0.677	?	0.222	?	0.677	?	?	?	0.677	?	0.286	?
	ROC	0.595	0.577	0.615	0.596	0.497	0.497	0.498	0.497	0.596	0.576	0.619	0.598
	Acuidad e	0.675				0.677				0.676			
T4		Naives Bayes				J48				Logistic			
U	Clases	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg

	Métricas												
S10	TP Rate	1	0	0	0.667	1	0	0	0.677	1	0	0	0.677
	Precisión	0.677	?	?	?	0.677	?	?	?	0.677	?	?	?
	ROC	0.531	0.549	0.501	0.528	0.497	0.497	0.498	0.497	0.532	0.551	0.499	0.529
	Acuidad e	0.677				0.677				0.677			
S11	TP Rate	1	0	0	0.667	1	0	0	0.677	1	0	0	0.677
	Precisión	0.677	?	?	?	0.677	?	?	?	0.677	?	?	?
	ROC	0.506	0.510	0.552	0.515	0.497	0.497	0.498	0.497	0.502	0.507	0.548	0.511
	Acuidad e	0.677				0.677				0.677			
S12	TP Rate	1	0	0	0.667	1	0	0	0.677	1	0	0	0.677
	Precisión	0.677	?	?	?	0.677	?	?	?	0.677	?	?	?
	ROC	0.530	0.499	0.527	0.525	0.497	0.497	0.498	0.497	0.530	0.499	0.527	0.525
	Acuidad e	0.677				0.677				0.677			
S13	TP Rate	1	0	0	0.667	1	0	0	0.677	1	0	0	0.677
	Precisión	0.677	?	?	?	0.677	?	?	?	0.677	?	?	?
	ROC	0.517	0.513	0.507	0.514	0.497	0.497	0.498	0.497	0.517	0.513	0.507	0.514
	Acuidad e	0.677				0.677				0.677			
S14	TP Rate	1	0	0	0.667	1	0	0	0.677	1	0	0	0.677
	Precisión	0.677	?	?	?	0.677	?	?	?	0.677	?	?	?
	ROC	0.539	0.509	0.534	0.34	0.497	0.497	0.498	0.497	0.539	0.509	0.534	0.534
	Acuidad e	0.677				0.677				0.677			
S15	TP Rate	1	0	0	0.667	1	0	0	0.677	1	0	0	0.677
	Precisión	0.677	?	?	?	0.677	?	?	?	0.677	?	?	?
	ROC	0.516	0.392	0.546	0.505	0.497	0.497	0.498	0.497	0.516	0.391	0.546	0.505
	Acuidad e	0.677				0.677				0.677			

Anexo IV

Tabela 32 – Resultados da análise aos dados do concessionário na gama Luxo/Supercarro para o Target T2

T2	Naives Bayes					J48					Logistic					
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S1	TP Rate	0	0.154	0	0.932	0.614	0	0	0	1	0.629	0	0.154	0.200	0.795	0.557
	Precision	0	0.400	?	0.641	?	?	?	?	0.629	?	0	0.400	0.286	0.673	0.538
	ROC	0.697	0.402	0.626	0.612	0.579	0.134	0.401	0.500	0.427	0.420	0.701	0.351	0.483	0.566	0.520
	Acuidade	0.614					0.629					0.557				
S2	TP Rate	0	0	0	0.995	0.600	0	0	0	1	0.629	0	0.077	0	0.932	0.600
	Precision	?	0	?	0.627	?	?	?	?	0.629	?	0	0.333	?	0.631	?
	ROC	0.366	0.401	0.591	0.508	0.494	0.134	0.401	0.500	0.427	0.420	0.316	0.425	0.498	0.403	0.417
	Acuidade	0.600					0.692					0.600				
S3	TP Rate	0	0	0	0.997	0.614	0	0	0	1	0.629	0	0	0	0.997	0.614
	Precision	?	0	0	0.662	?	?	?	?	0.629	?	?	?	0	0.642	?
	ROC	0.701	0.441	0.477	0.611	0.564	0.134	0.401	0.500	0.427	0.420	0.612	0.374	0.488	0.599	0.542
	Acuidade	0.641					0.629					0.614				
S4	TP Rate	0	0	0	0.995	0.600	0	0	0	1	0.629	0	0	0	0.909	0.571
	Precision	0	?	0	0.636	?	?	?	?	0.629	?	0	0	0	0.635	0.399
	ROC	0.627	0.422	0.439	0.577	0.531	0.134	0.401	0.500	0.427	0.420	0.701	0.448	0.461	0.598	0.555
	Acuidade	0.600					0.692					0.571				
T2	Naives Bayes					J48					Logistic					
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S5	TP Rate	0	0.077	0	0.932	0.600	0	0	0	1	0.629	0	0.077	0.100	0.864	0.571
	Precision	?	0.200	0	0.641	?	?	?	?	0.629	?	0	0.333	0.333	0.623	0.501
	ROC	0.776	0.441	0.689	0.654	0.625	0.134	0.401	0.500	0.427	0.420	0.677	0.422	0.503	0.528	0.511
	Acuidade	0.600					0.629					0.571				
S6	TP Rate	0	0.077	0	0.909	0.586	0	0	0	1	0.629	0	0.231	0.300	0.841	0.614
	Precision	0	0.250	0	0.625	0.439	?	?	?	0.629	?	0	0.600	0.375	0.673	0.588
	ROC	0.749	0.431	0.640	0.611	0.588	0.134	0.401	0.500	0.427	0.420	0.803	0.447	0.515	0.538	0.529

	Acuidade	0.586					0.629					0.614				
S7	TP Rate	0	0.154	0	0.955	0.629	0	0	0	1	0.629	0	0.154	0.100	0.841	0.571
	Precision	0	0.500	?	0.646	?	?	?	?	0.629	?	0	0.400	0.250	0.661	0.515
	ROC	0.682	0.361	0.635	0.594	0.560	0.134	0.401	0.500	0.427	0.420	0.468	0.353	0.508	0.519	0.484
	Acuidade	0.629					0.629					0.571				
S8	TP Rate	0	0.154	0	0.932	0.614	0	0	0	1	0.629	0	0.154	0.200	0.795	0.557
	Precision	0	0.400	?	0.641	?	?	?	?	0.629	?	0	0.400	0.286	0.673	0.538
	ROC	0.697	0.402	0.626	0.612	0.579	0.134	0.401	0.500	0.427	0.420	0.701	0.351	0.483	0.566	0.520
	Acuidade	0.614					0.629					0.557				
S9	TP Rate	0	0.154	0	0.955	0.509	0	0	0	1	0.629	0	0.154	0	0.818	0.543
	Precision	?	0.400	?	0.646	?	?	?	?	0.629	?	0	0.333	0	0.610	0.445
	ROC	0.413	0.414	0.589	0.525	0.509	0.134	0.401	0.500	0.427	0.420	0.323	0.425	0.484	0.484	0.466
	Acuidade	0.509					0.629					0.543				
T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0	0	0	0.955	0.600	0	0	0	1	0.629	0	0	0	0.995	0.600
	Precision	?	0	?	0.618	?	?	?	?	0.629	?	?	0	?	0.618	?
	ROC	0.517	0.441	0.662	0.549	0.544	0.134	0.401	0.500	0.427	0.420	0.403	0.448	0.527	0.432	0.447
	Acuidade	0.600					0.629					0.600				
S11	TP Rate	0	0	0	1	0.629	0	0	0	1	0.629	0	0	0	1	0.629
	Precision	?	?	?	0.629	?	?	?	?	0.629	?	?	?	?	0.629	?
	ROC	0.070	0.398	0.228	0.385	0.351	0.134	0.401	0.500	0.427	0.420	0.045	0.432	0.348	0.374	0.367
	Acuidade	0.629					0.629					0.629				
S12	TP Rate	0	0	0	1	0.629	0	0	0	1	0.629	0	0	0	1	0.629
	Precision	?	?	?	0.629	?	?	?	?	0.629	?	?	?	?	0.629	?
	ROC	0.102	0.425	0.404	0.468	0.435	0.134	0.401	0.500	0.427	0.420	0.062	0.345	0.371	0.470	0.415
	Acuidade	0.629					0.629					0.629				
S13	TP Rate	0	0	0	1	0.629	0	0	0	1	0.629	0	0	0	1	0.629
	Precision	?	?	?	0.629	?	?	?	?	0.629	?	?	?	?	0.629	?
	ROC	0.391	0.520	0.283	0.479	0.455	0.134	0.401	0.500	0.427	0.420	0.391	0.520	0.269	0.479	0.453

	Acuidade	0.629					0.629					0.629				
S14	TP Rate	0	0	0	1	0.629	0	0	0	1	0.629	0	0	0	1	0.629
	Precision	?	?	?	0.629	?	?	?	?	0.629	?	?	?	?	0.629	?
	ROC	0.134	0.401	0.500	0.427	0.420	0.134	0.401	0.500	0.427	0.420	0.134	0.401	0.500	0.427	0.420
	Acuidade	0.629					0.629					0.629				
S15	TP Rate	0	0	0	1	0.629	0	0	0	1	0.629	0	0	0	1	0.629
	Precision	?	?	?	0.629	?	?	?	?	0.629	?	?	?	?	0.629	?
	ROC	0.669	0.292	0.487	0.528	0.484	0.134	0.401	0.500	0.427	0.420	0.669	0.248	0.487	0.525	0.474
	Acuidade	0.629					0.629					0.629				

Tabela 33 – Resultados da análise aos dados do concessionário na gama Luxo/Supercarro para o Target T3

T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S1	TP Rate	0	0.429	0	0.750	0.514	0	0	0	0.889	0.457	0	0.429	0	0.778	0.529
	Precision	0	0.450	0	0.563	0.424	?	0	0	0.492	?	0	0.529	0	0.583	0.459
	ROC	0.426	0.624	0.368	0.581	0.559	0.300	0.501	0.334	0.429	0.431	0.297	0.611	0.445	0.564	0.545
	Acuidade	0.514					0.457					0.529				
S2	TP Rate	0	0.190	0	0.806	0.471	0	0	0	0.944	0.486	0	0.143	0	0.778	0.443
	Precision	?	0.308	?	0.509	?	?	0	?	0.507	?	?	0.250	0	0.500	?
	ROC	0.297	0.558	0.527	0.529	0.521	0.338	0.532	0.433	0.453	0.466	0.303	0.561	0.608	0.501	0.517
	Acuidade	0.471					0.486					0.443				
S3	TP Rate	0	0.286	0	0.806	0.500	0	0.095	0	0.861	0.471	0	0.286	0	0.806	0.500
	Precision	?	0.375	?	0.537	?	?	0.222	?	0.508	?	?	0.375	?	0.537	?
	ROC	0.528	0.576	0.310	0.554	0.531	0.295	0.473	0.403	0.472	0.452	0.543	0.552	0.433	0.527	0.525
	Acuidade	0.500					0.471					0.500				
S4	TP Rate	0	0.286	0	0.722	0.457	0	0.095	0	0.861	0.471	0	0.381	0	0.722	0.486
	Precision	?	0.316	?	0.510	?	?	0.222	?	0.508	?	?	0.364	?	0.542	?
	ROC	0.405	0.580	0.279	0.503	0.494	0.362	0.518	0.343	0.476	0.465	0.408	0.609	0.352	0.542	0.531
	Acuidade	0.457					0.471					0.486				
T3		Naives Bayes					J48					Logistic				
Ce	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg

	Métricas															
S5	TP Rate	0	0.429	0	0.806	0.543	0	0	0	0.861	0.443	0	0.429	0	0.806	0.543
	Precision	?	0.450	0	0.806	?	?	0	0	0.484	?	0	0.529	0	0.569	0.451
	ROC	0.534	0.625	0.371	0.624	0.589	0.338	0.459	0.298	0.447	0.426	0.417	0.626	0.486	0.575	0.569
	Acuidade	0.543					0.443					0.543				
S6	TP Rate	0	0.381	0	0.778	0.514	0	0	0	0.889	0.457	0	0.429	0	0.806	0.543
	Precision	0	0.471	0	0.549	0.424	?	0	?	0.492	?	0	0.500	0	0.592	0.560
	ROC	0.494	0.603	0.416	0.581	0.563	0.300	0.494	0.388	0.454	0.447	0.429	0.610	0.507	0.560	0.560
	Acuidade	0.514					0.457					0.543				
S7	TP Rate	0	0.429	0	0.778	0.529	0	0	0	0.917	0.529	0	0.476	0.125	0.778	0.557
	Precision	?	0.500	0	0.549	?	?	0	0	0.500	?	?	0.526	0.500	0.571	?
	ROC	0.320	0.645	0.426	0.572	0.560	0.354	0.517	0.418	0.449	0.459	0.160	0.644	0.509	0.516	0.528
	Acuidade	0.529					0.529					0.557				
S8	TP Rate	0	0.429	0	0.750	0.514	0	0	0	0.889	0.457	0	0	0	1	0.514
	Precision	0	0.450	0	0.563	0.424	?	0	0	0.492	?	?	?	?	0.514	?
	ROC	0.426	0.624	0.368	0.581	0.559	0.300	0.501	0.334	0.429	0.431	0.231	0.469	0.387	0.431	0.423
	Acuidade	0.514					0.457					0.514				
S9	TP Rate	0	0.286	0	0.889	0.543	0	0	0	0.917	0.471	0	0.333	0	0.722	0.471
	Precision	?	0.545	0	0.552	?	?	0	0	0.500	?	?	0.350	?	0.520	?
	ROC	0.600	0.569	0.436	0.543	0.525	0.286	0.518	0.353	0.433	0.439	0.457	0.512	0.421	0.502	0.492
	Acuidade	0.543					0.471					0.471				
T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0	0.143	0	0.917	0.514	0	0	0	0.972	0.500	0	0.095	0	0.917	0.500
	Precision	?	0.375	?	0.532	?	?	0	?	0.507	?	?	0.286	0	0.541	?
	ROC	0.362	0.534	0.519	0.557	0.532	0.277	0.462	0.343	0.402	0.404	0.358	0.582	0.580	0.502	0.524
	Acuidade	0.514					0.500					0.500				
S11	TP Rate	0	0	0	1	0.514	0	0	0	1	0.514	0	0.048	0	0.917	0.486
	Precision	?	?	?	0.514	?	?	?	?	0.514	?	?	0.250	?	0.500	?
	ROC	0.129	0.461	0.374	0.291	0.340	0.231	0.469	0.387	0.431	0.423	0.154	0.463	0.402	0.304	0.352

	Acuidade	0.514					0.514					0.486				
S12	TP Rate	0	0	0	0.972	0.500	0	0	0	1	0.514	0	0.048	0	0.994	0.500
	Precision	?	0	?	0.507	?	?	?	?	0.514	?	?	0.250	?	0.515	?
	ROC	0.134	0.500	0.419	0.437	0.432	0.231	0.469	0.387	0.431	0.423	0.134	0.500	0.419	0.434	0.431
	Acuidade	0.500					0.514					0.500				
S13	TP Rate	0	0	0	1	0.514	0	0	0	1	0.514	0	0	0	1	0.514
	Precision	?	?	?	0.514	?	?	?	?	0.514	?	?	?	?	0.514	?
	ROC	0.445	0.435	0.171	0.438	0.430	0.231	0.469	0.387	0.431	0.423	0.445	0.435	0.228	0.438	0.437
	Acuidade	0.514					0.514					0.514				
S14	TP Rate	0	0	0	1	0.514	0	0	0	1	0.514	0	0	0	1	0.514
	Precision	?	?	?	0.514	?	?	?	?	0.514	?	?	?	?	0.514	?
	ROC	0.231	0.469	0.387	0.431	0.423	0.231	0.469	0.387	0.431	0.423	0.231	0.469	0.387	0.431	0.423
	Acuidade	0.514					0.514					0.514				
S15	TP Rate	0	0.286	0	0.722	0.457	0	0.143	0	0.806	0.457	0	0.333	0	0.722	0.492
	Precision	?	0.316	?	0.510	?	?	0.250	?	0.500	?	?	0.350	?	0.520	?
	ROC	0.457	0.512	0.421	0.502	0.492	0.515	0.459	0.459	0.470	0.469	0.457	0.512	0.421	0.502	0.492
	Acuidade	0.457					0.457					0.492				

Anexo V

Tabela 34 – Resultados da análise aos dados do concessionário na gama SUV/Jipe para o Target T2

T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S1	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	0.995	0.491
	Precision	?	?	0	0.494	?	?	?	?	0.493	?	?	0	0	0.493	?
	ROC	0.532	0.518	0.461	0.511	0.505	0.485	0.491	0.492	0.497	0.494	0.524	0.525	0.453	0.522	0.510
	Acuidade	0.493					0.493					0.491				
S2	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.491	0.532	0.473	0.512	0.507	0.485	0.491	0.492	0.497	0.494	0.494	0.532	0.469	0.516	0.509
	Acuidade	0.493					0.493					0.493				
S3	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.540	0.450	0.472	0.490	0.482	0.485	0.491	0.492	0.497	0.494	0.525	0.450	0.472	0.495	0.484
	Acuidade	0.493					0.493					0.493				
S4	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.527	0.520	0.460	0.519	0.509	0.485	0.491	0.492	0.497	0.494	0.528	0.527	0.458	0.526	0.513
	Acuidade	0.493					0.493					0.493				
S5	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.548	0.475	0.469	0.493	0.490	0.485	0.491	0.492	0.497	0.494	0.522	0.480	0.469	0.503	0.493
	Acuidade	0.493					0.493					0.493				
T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S6	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.531	0.518	0.459	0.506	0.502	0.485	0.491	0.492	0.497	0.494	0.527	0.524	0.452	0.515	0.506

	Acuidade	0.493					0.493					0.493				
S7	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.523	0.524	0.462	0.511	0.506	0.485	0.491	0.492	0.497	0.494	0.522	0.532	0.451	0.516	0.508
	Acuidade	0.493					0.493					0.493				
S8	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.494	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.528	0.524	0.469	0.516	0.510	0.485	0.491	0.492	0.497	0.494	0.528	0.529	0.464	0.528	0.516
	Acuidade	0.493					0.493					0.493				
S9	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.531	0.519	0.465	0.515	0.508	0.485	0.491	0.492	0.497	0.494	0.521	0.524	0.460	0.526	0.513
	Acuidade	0.493					0.493					0.493				
T2		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.465	0.500	0.482	0.487	0.487	0.485	0.491	0.492	0.497	0.494	0.471	0.502	0.481	0.494	0.491
	Acuidade	0.493					0.493					0.493				
S11	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.512	0.523	0.472	0.494	0.498	0.485	0.491	0.492	0.497	0.494	0.511	0.523	0.472	0.491	0.496
	Acuidade	0.493					0.493					0.493				
S12	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.479	0.482	0.491	0.498	0.491	0.485	0.491	0.492	0.497	0.494	0.479	0.482	0.491	0.498	0.491
	Acuidade	0.493					0.493					0.493				
S13	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precision	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.485	0.441	0.487	0.490	0.478	0.485	0.491	0.492	0.497	0.494	0.485	0.442	0.486	0.490	0.478

	Acuidade	0.493					0.493					0.493				
S14	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precis ion	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.493	0.437	0.436	0.468	0.458	0.485	0.491	0.492	0.497	0.494	0.493	0.437	0.436	0.468	0.458
	Acuidade	0.493					0.493					0.493				
S15	TP Rate	0	0	0	1	0.493	0	0	0	1	0.493	0	0	0	1	0.493
	Precis ion	?	?	?	0.493	?	?	?	?	0.493	?	?	?	?	0.493	?
	ROC	0.507	0.462	0.475	0.469	0.472	0.485	0.491	0.492	0.497	0.494	0.507	0.462	0.475	0.470	0.472
	Acuidade	0.493					0.493					0.493				

Tabela 35 – Resultados da análise aos dados do concessionário na gama SUV/Jipe para o Target T3

T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S1	TP Rate	0.011	0.343	0	0.612	0.311	0.034	0.408	0.016	0.648	0.350	0.023	0.339	0.016	0.641	0.324
	Precis ion	0.105	0.282	0	0.340	0.211	0.182	0.330	0.125	0.382	0.284	0.143	0.280	0.250	0.362	0.275
	ROC	0.517	0.492	0.546	0.532	0.519	0.548	0.524	0.476	0.578	0.540	0.515	0.486	0.538	0.543	0.520
	Acuidade	0.311					0.350					0.324				
S2	TP Rate	0	0.424	0	0.662	0.350	0	0	0	1	0.339	0	0.437	0	0.655	0.351
	Precis ion	?	0.335	0	0.359	?	?	?	?	0.339	?	?	0.331	0	0.364	0.509
	ROC	0.494	0.501	0.507	0.526	0.510	0.491	0.493	0.497	0.497	0.494	0.489	0.503	0.506	0.529	0.509
	Acuidade	0.350					0.339					0.351				
S3	TP Rate	0	0.296	0	0.733	0.328	0	0.343	0	0.794	0.370	0	0.306	0	0.701	0.328
	Precis ion	?	0.275	?	0.350	?	0	0.353	?	0.384	?	?	0.273	?	0.356	?
	ROC	0.506	0.478	0.552	0.531	0.513	0.538	0.514	0.545	0.564	0.541	0.501	0.472	0.557	0.525	0.509
	Acuidade	0.328					0.370					0.328				
S4	TP Rate	0	0.355	0	0.690	0.339	0	0.424	0	0.705	0.364	0	0.363	0	0.690	0.341
	Precis ion	0	0.302	?	0.359	?	0	0.384	0	0.378	0.241	0	0.295	?	0.370	?
	ROC	0.518	0.496	0.557	0.542	0.526	0.474	0.561	0.508	0.563	0.535	0.505	0.490	0.555	0.545	0.522
	Acuidade	0.339					0.364					0.341				
T3		Naives Bayes					J48					Logistic				
Ce	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg

	Métricas															
S5	TP Rate	0	0.355	0	0.683	0.331	0.011	0.314	0	0.733	0.344	0.011	0.306	0.008	0.705	0.333
	Precision	0	0.287	0	0.355	0.205	0.118	0.325	0	0.365	0.245	0.100	0.287	0.250	0.364	0.509
	ROC	0.507	0.476	0.546	0.520	0.508	0.530	0.519	0.525	0.538	0.529	0.510	0.471	0.543	0.527	0.509
	Acuidade	0.331					0.344					0.333				
S6	TP Rate	0.006	0	0.351	0.651	0.326	0.057	0.367	0.016	0.665	0.349	0.023	0.363	0.008	0.655	0.335
	Precision	0.056	0.299	0	0.352	0.219	0.233	0.331	0.111	0.377	0.292	0.143	0.306	0.125	0.367	0.264
	ROC	0.518	0.494	0.549	0.524	0.518	0.556	0.527	0.474	0.578	0.542	0.512	0.488	0.543	0.537	0.518
	Acuidade	0.326					0.349					0.335				
S7	TP Rate	0	0.363	0	0.630	0.321	0.052	0.392	0.023	0.527	0.309	0	0.376	0	0.633	0.326
	Precision	0	0.289	0	0.346	0.203	0.056	0.428	0.047	0.662	0.261	0	0.293	0	0.350	0.205
	ROC	0.495	0.499	0.550	0.538	0.519	0.530	0.510	0.474	0.548	0.521	0.493	0.495	0.538	0.542	0.517
	Acuidade	0.321					0.309					0.326				
S8	TP Rate	0	0.367	0	0.641	0.326	0.011	0.437	0.047	0.557	0.334	0.034	0.343	0	0.644	0.327
	Precision	0	0.304	0	0.347	0.208	0.063	0.313	0.207	0.380	0.267	0.214	0.290	0	0.355	0.251
	ROC	0.523	0.498	0.524	0.540	0.522	0.507	0.534	0.495	0.552	0.528	0.523	0.493	0.521	0.548	0.522
	Acuidade	0.326					0.334					0.327				
S9	TP Rate	0.011	0.367	0	0.641	0.328	0.011	0.392	0	0.740	0.369	0.034	0.355	0	0.673	0.340
	Precision	0.125	0.294	0	0.359	0.235	0.182	0	0.360	0.373	0.288	0.100	0.374	0	0.374	0.255
	ROC	0.519	0.490	0.530	0.534	0.515	0.512	0.528	0.479	0.572	0.532	0.520	0.486	0.528	0.538	0.517
	Acuidade	0.328					0.369					0.340				
T3		Naives Bayes					J48					Logistic				
Cenários	Classes	A	B	C	D	Avg	A	B	C	D	Avg	A	B	C	D	Avg
	Métricas															
S10	TP Rate	0	0.363	0	0.651	0.328	0	0	0	1	0.339	0	0.384	0	0.633	0.328
	Precision	?	0.303	?	0.342	?	?	?	?	0.339	?	?	0.300	?	0.345	?
	ROC	0.484	0.489	0.476	0.497	0.486	0.491	0.493	0.497	0.497	0.494	0.481	0.480	0.471	0.501	0.486
	Acuidade	0.328					0.339					0.328				
S11	TP Rate	0	0.224	0	0.882	0.345	0	0	0	1	0.339	0	0.224	0	0.822	0.345
	Precision	?	0.314	?	0.353	?	?	?	?	0.339	?	?	0.314	?	0.353	?
	ROC	0.468	0.496	0.520	0.520	0.502	0.491	0.493	0.497	0.497	0.494	0.468	0.498	0.521	0.520	0.503

	Acuidade	0.345					0.339					0.345				
S12	TP Rate	0	0	0	1	0.339	0	0	0	1	0.339	0	0	0	1	0.339
	Precis ion	?	?	?	0.339	?	?	?	?	0.339	?	?	?	?	0.339	?
	ROC	0.487	0.496	0.470	0.509	0.495	0.491	0.493	0.497	0.497	0.494	0.487	0.496	0.472	0.509	0.495
	Acuidade	0.339					0.339					0.339				
S13	TP Rate	0	0	0	1	0.339	0	0	0	1	0.339	0	0	0	1	0.339
	Precis ion	?	?	?	0.339	?	?	?	?	0.339	?	?	?	?	0.339	?
	ROC	0.497	0.454	0.447	0.479	0.470	0.491	0.493	0.497	0.497	0.494	0.497	0.454	0.450	0.479	0.471
	Acuidade	0.339					0.339					0.339				
S14	TP Rate	0	0	0	1	0.339	0	0	0	1	0.339	0	0	0	1	0.339
	Precis ion	?	?	?	0.339	?	?	?	?	0.339	?	?	?	?	0.339	?
	ROC	0.474	0.437	0.534	0.482	0.475	0.491	0.493	0.497	0.497	0.494	0.437	0.474	0.534	0.482	0.475
	Acuidade	0.339					0.339					0.339				
S15	TP Rate	0	0	0	1	0.339	0	0	0	1	0.339	0	0	0	1	0.339
	Precis ion	?	?	?	0.339	?	?	?	?	0.339	?	?	?	?	0.339	?
	ROC	0.487	0.494	0.525	0.515	0.504	0.491	0.493	0.497	0.497	0.494	0.487	0.494	0.525	0.515	0.504
	Acuidade	0.339					0.339					0.339				

Tabela 36 – Resultados da análise aos dados do concessionário na gama SUV/Jipe para o Target T4

T4		Naives Bayes				J48				Logistic			
Cenário	Classes	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
S1	TP Rate	0.993	0	0	0.656	1	0	0	0.661	0.995	0	0	0.657
	Precis ion	0.660	?	0	?	0.661	?	?	?	0.660	?	0	?
	ROC	0.549	0.519	0.584	0.550	0.497	0.491	0.493	0.495	0.549	0.527	0.578	0.549
	Acuidade	0.656				0.661				0.657			
S2	TP Rate	0.995	0	0	0.657	1	0	0	0.661	0.996	0	0	0.659
	Precis ion	0.660	?	0	?	0.661	?	?	?	0.660	?	0	?
	ROC	0.522	0.434	0.587	0.516	0.497	0.491	0.493	0.495	0.517	0.434	0.577	0.511
	Acuidade	0.657				0.661				0.659			
S3	TP Rate	1	0	0	0.661	1	0	0	0.661	1	0	0	0.661
	Precis ion	0.661	?	?	?	0.661	?	?	?	0.661	?	?	?

	ROC	0.546	0.551	0.520	0.543	0.497	0.491	0.493	0.495	0.539	0.555	0.517	0.538
	Acuidad e	0.661				0.661				0.661			
S4	TP Rate	1	0	0	0.661	1	0	0	0.661	1	0	0	0.661
	Precisio n	0.661	?	?	?	0.661	?	?	?	0.661	?	?	?
	ROC	0.557	0.550	0.582	0.559	0.497	0.491	0.493	0.495	0.555	0.550	0.575	0.557
	Acuidad e	0.661				0.661				0.661			
T4		Naives Bayes				J48				Logistic			
Cenários	Clases	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg
	Métricas												
S5	TP Rate	0.998	0	0	0.660	1	0	0	0.661	0.998	0	0	0.660
	Precisio n	0.661	?	0	?	0.661	?	?	?	0.661	?	0	?
	ROC	0.532	0.515	0.522	0.527	0.497	0.491	0.493	0.495	0.528	0.519	0.510	0.523
	Acuidad e	0.660				0.661				0.660			
S6	TP Rate	0.993	0	0	0.656	1	0	0	0.661	0.995	0	0	0.657
	Precisio n	0.661	?	0	?	0.661	?	?	?	0.660	?	0	?
	ROC	0.542	0.495	0.590	0.541	0.497	0.491	0.493	0.495	0.540	0.504	0.583	0.540
	Acuidad e	0.656				0.661				0.657			
S7	TP Rate	0.995	0	0	0.657	1	0	0	0.661	0.995	0	0	0.657
	Precisio n	0.660	?	0	?	0.661	?	?	?	0.660	?	0	?
	ROC	0.547	0.487	0.592	0.543	0.497	0.491	0.493	0.495	0.539	0.474	0.583	0.534
	Acuidad e	0.657				0.661				0.657			
S8	TP Rate	0.993	0	0	0.656	1	0	0	0.661	0.995	0	0	0.657
	Precisio n	0.660	?	0	?	0.661	?	?	?	0.660	?	0	?
	ROC	0.548	0.523	0.591	0.550	0.497	0.491	0.493	0.495	0.549	0.530	0.582	0.551
	Acuidad e	0.656				0.661				0.657			
S9	TP Rate	0.995	0	0	0.657	1	0	0	0.661	0.995	0	0	0.657
	Precisio n	0.660	?	0	?	0.661	?	?	?	0.660	?	0	?
	ROC	0.545	0.525	0.581	0.547	0.497	0.491	0.493	0.495	0.543	0.533	0.571	0.545
	Acuidad e	0.657				0.661				0.657			
T4		Naives Bayes				J48				Logistic			
C	Clases	A	B	C	Avg	A	B	C	Avg	A	B	C	Avg

	Métricas												
S10	TP Rate	1	0	0	0.661	1	0	0	0.661	1	0	0	0.661
	Precisión	0.661	?	?	?	0.661	?	?	?	0.661	?	?	?
	ROC	0.493	0.335	0.506	0.469	0.497	0.491	0.493	0.495	0.489	0.356	0.502	0.467
	Acuidad	0.661				0.661				0.661			
S11	TP Rate	1	0	0	0.661	1	0	0	0.661	1	0	0	0.661
	Precisión	0.661	?	?	?	0.661	?	?	?	0.661	?	?	?
	ROC	0.516	0.480	0.540	0.513	0.497	0.491	0.493	0.495	0.516	0.483	0.545	0.514
	Acuidad	0.661				0.661				0.661			
S12	TP Rate	1	0	0	0.661	1	0	0	0.661	1	0	0	0.661
	Precisión	0.661	?	?	?	0.661	?	?	?	0.661	?	?	?
	ROC	0.505	0.516	0.463	0.500	0.497	0.491	0.493	0.495	0.505	0.516	0.471	0.502
	Acuidad	0.661				0.661				0.661			
S13	TP Rate	1	0	0	0.661	1	0	0	0.661	1	0	0	0.661
	Precisión	0.661	?	?	?	0.661	?	?	?	0.661	?	?	?
	ROC	0.492	0.509	0.480	0.493	0.497	0.491	0.493	0.495	0.492	0.509	0.480	0.493
	Acuidad	0.661				0.661				0.661			
S14	TP Rate	1	0	0	0.661	1	0	0	0.661	1	0	0	0.661
	Precisión	0.661	?	?	?	0.661	?	?	?	0.661	?	?	?
	ROC	0.501	0.478	0.480	0.493	0.497	0.491	0.493	0.495	0.501	0.478	0.480	0.493
	Acuidad	0.661				0.661				0.661			
S15	TP Rate	1	0	0	0.661	1	0	0	0.661	1	0	0	0.661
	Precisión	0.661	?	?	?	0.661	?	?	?	0.661	?	?	?
	ROC	0.509	0.469	0.523	0.504	0.497	0.491	0.493	0.495	0.509	0.469	0.523	0.504
	Acuidad	0.661				0.661				0.661			