

An Intelligent Model to Assist Customers to Choose New Vehicles

F. A. S. Vince, L. de C. Nogueira and R. M. Salgado

Abstract— The auto industry invests increasingly in offering various car models in order to conquer the consumer market. Due to the major portfolio of options on the market and the different consumption profiles existence, identify the motivational factors when we need purchasing a new vehicle is not a simple task. Emotional and rational factors are included in the knowledge and judgment of the people on whom or what features are relevant in decision buying process of an automobile. From this viewpoint, this paper proposes the development of an application that can assist in choosing a new car that meets the user desires. We combine, an expert system with clusters algorithm to develop an intelligent model to help customers in hard decision of choose a car. In terms of results, we can say that the proposed model was able to show excellent options to customers, associating car models with the desired characteristics of each customer.

Keywords— Smart Suggestions, Fuzzy C-Means, Subtractive Clustering, Consumer Profiles, Vehicles Suggestion.

I. INTRODUÇÃO

OS SERES humanos possuem diversos tipos de necessidades, sejam estas de segurança, fisiológicas, relacionamentos, de estima ou autorrealização. O consumo está intimamente ligado à missão de suprir as necessidades humanas. O ato de consumir é a atividade presente na aquisição de bens e serviços pelos indivíduos ou empresas. Há muito tempo o homem possui a necessidade de adquirir bens de consumo, dos mais simples relacionados à alimentação ou higiene aos mais complexos como automóveis, imóveis e outros bens duráveis.

A escolha de um produto envolve uma sequência típica de eventos, chamada de processo de compra: reconhecimento do problema, busca de informações, avaliação de alternativas, decisão de compra e comportamento pós-compra. O número de etapas percorridas e o tempo consumido em cada etapa variam entre indivíduos, produtos e circunstâncias de compra [1].

Considerando a compra de um automóvel, este processo de compra pode ser motivado por necessidade de deslocamento; transporte de familiares, carga; seu uso pode ser voltado para lazer ou trabalho, e o consumidor pode priorizar algumas características do veículo como potência do motor, capacidade do porta-malas, e até mesmo subjetivas como o *design* da carroceria.

O automóvel na sociedade brasileira não é apenas um bem durável para locomoção, mas também um objeto de *status* e sua conquista, uma autorrealização pessoal.

O mercado automobilístico brasileiro é um dos maiores do mundo, ocupando em 2014, com a marca de 3.328.958 veículos, a 4ª posição em número de vendas anuais de veículos novos no ranking mundial.

Segundo o anuário da indústria brasileira de veículos edição 2015, atualizado até o mês de maio e publicado pela Associação Nacional dos Fabricantes de Veículos Automotores (Anfavea) [2], o mercado automobilístico nacional possui atualmente 64 unidades industriais espalhadas por 52 municípios em 10 estados. Esta indústria representa 23% do produto interno bruto (PIB) industrial e 5% do PIB total, com faturamento acima de US\$110 bilhões. Além destes dados, também se ressalta a oferta de 1,5 milhão de empregos direta e indiretamente, somando-se ainda uma arrecadação de carga tributária no valor de US\$178,5 bilhões em 2013.

Neste contexto, percebe-se o grau de importância desta indústria de ritmo crescente em nossa sociedade. O consumidor final é a peça chave para o ideal funcionamento deste setor e a indústria automobilística opera a fim de conquistar a preferência deste para investir em seus produtos.

Os seres humanos possuem como uma de suas características a individualidade, presente não apenas em fatores biológicos, mas também em tomada de decisões, opiniões e preferências.

Identificar os fatores motivacionais na aquisição de um novo veículo não é uma tarefa simples: fatores emocionais e racionais se misturam no conhecimento e julgamento das pessoas sobre qual ou quais características são relevantes no processo decisório de compra.

Dada a relevância do tema abordado, a metodologia deste trabalho propõe a criação de uma aplicação capaz de sugerir automóveis condizentes com as características consideradas relevantes pelo consumidor de um novo veículo. Esta sugestão será baseada em perfis de consumidores que sejam semelhantes ao do novo consumidor.

Este artigo está organizado da seguinte forma: na seção II são apresentados os trabalhos relacionados, já na seção III é apresentada a fundamentação teórica das técnicas utilizadas neste trabalho, na seção IV é exposta a metodologia utilizada para desenvolvimento da pesquisa, na seção V são descritas as bases de dados utilizadas neste trabalho, na seção VI é mostrado os resultados obtidos pelos agrupamentos e finalmente na seção VII são feitas as considerações finais.

F. A. S. Vince, Universidade Federal de Alfenas (UNIFAL-MG), Alfenas, Minas Gerais, Brasil, a08008@bcc.unifal-mg.edu.br

L. de C. Nogueira, Universidade Federal de Alfenas (UNIFAL-MG), Alfenas, Minas Gerais, Brasil, leandra.carvalho@bcc.unifal-mg.edu.br

R. M. Salgado, Universidade Federal de Alfenas (UNIFAL-MG), Alfenas, Minas Gerais, Brasil, ricardo@bcc.unifal-mg.edu.br

II. TRABALHOS RELACIONADOS

Na literatura existem diversos trabalhos que visam classificar automaticamente conjuntos de dados. Dentre os algoritmos utilizados pelos autores dos trabalhos apresentados se encontram o FCM e SC, que também são utilizados neste trabalho.

No trabalho de [3] foi realizado um estudo combinando os algoritmos de classificação FCM e SC. Na simulação realizada, utilizou-se como entrada para o SC uma base de dados sintética composta por 300 pontos de dados gerados aleatoriamente, e o resultado do SC foi reaproveitado como número de grupos pelo FCM. Chegou-se a conclusão que na maioria dos casos, a combinação de técnicas apresentou melhor taxa de convergência e menor número de iterações em relação a utilizar apenas o FCM, além de dispensar a necessidade de informar o número de grupos previamente para o FCM.

O trabalho de [4] apresentou a investigação sobre a utilização de fontes de recomendação fortes e fracas na escolha de veículos, onde se realizou um experimento disponibilizado na internet utilizando a técnica de *Procces-Assisted Choice Modeling* (PACMod) combinado com questionários. Com nível de significância de 5%, algumas hipóteses mostraram-se significativas: quanto maior a importância das informações afetivas e maior o nível de dificuldade percebido na escolha, mais provável será a busca por recomendação de fontes fortes, e quanto maior a experiência prévia sobre o produto, maior serão os conhecimentos prévios objetivos e subjetivos. Também se verificou que os construtos “nível de dificuldade” e “tipos de informações” mostraram significativa ligação com as fontes de recomendações escolhidas.

No trabalho de [5], foram utilizadas técnicas de agrupamento para agregar os barramentos, na aplicação de modelos de previsão de séries temporais a fim de prever o consumo de carga diária ativa nos barramentos do sistema elétrico. As técnicas utilizadas para agrupar barramentos com perfis de consumo similares foram FCM e *Pruning Technique Kohonen Self-Organizing Feature Map* (PSOM). Ambas apresentaram resultados similares e eficientes, os agrupamentos obtidos foram pouco divergentes entre as duas técnicas.

No trabalho de [6], a autora tentou identificar os tipos de influência em que grupos de referência agem sobre o consumidor de veículos utilitários esportivos de tração 4x4. Utilizando questionários eletrônicos auto-aplicáveis e estruturados a uma base de 2000 clientes da fabricante de utilitários estudada, concluiu-se que as três formas de influência mais relevantes foram influência informacional, influência utilitária e influência expressiva de valor, respectivamente.

De acordo com o trabalho de [7], foi buscado o agrupamento de trabalhadores com perfis de aprendizado similares, a fim de evitar gargalos e perda de produtividade nas linhas de produção da indústria calçadista. Foi realizada uma seleção de 21 trabalhadores e coleta de dados dos tempos de execução das tarefas. Posteriormente uma obtenção dos parâmetros para cada modelo de curva de aprendizado foi realizada, seguida de um agrupamento dos trabalhadores pelo FCM. Foi verificado que a transformação nos dados originais promoveu um maior impacto nos resultados do que a técnica utilizada para clusterização.

No trabalho de [8], foram estabelecidos objetivos para identificar informações que são relevantes para um melhor entendimento sobre o comportamento de compra do consumidor de automóveis novos. Através de entrevistas pessoais com uso de questionários previamente testados, em diferentes concessionárias autorizadas da mesma cidade, concluiu-se que mais de 70% comprariam um novo veículo da mesma marca, há preocupação com a aceitação social da escolha, e as classes C, D e E estão comprando mais veículos, graças a incentivos financeiros. Também foi apontado que as esposas têm participado mais ativamente no processo decisório da compra, e fatores econômicos como preço de compra são de grande importância, ainda que suplantados pela qualidade.

No trabalho [9] foram realizadas entrevistas com jovens entre 18 e 24 anos utilizando a técnica de *laddering*, que possuíam um veículo próprio de passeio e participaram do processo de escolha no momento da compra, buscando compreender quais os valores que motivam a compra de um veículo novo. Os resultados demonstraram que os principais valores foram bem-estar, segurança, satisfação, preocupação com o futuro (preço, valor de revenda, valorização da marca) e qualidade de vida.

III. FUNDAMENTAÇÃO TEÓRICA

Antes de apresentar a metodologia abordada, serão apresentados os aspectos teóricos utilizados para a execução e desenvolvimento da pesquisa.

A. Classificação e Agrupamento

Um conjunto de dados D é definido como mostrado na Fig. 1.

$$D = \begin{bmatrix} \vec{d}_1 \\ \vec{d}_2 \\ \dots \\ \vec{d}_n \end{bmatrix} = \begin{bmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,m} \\ d_{2,1} & d_{2,2} & \dots & d_{2,m} \\ \dots & \dots & \dots & \dots \\ d_{n,1} & d_{n,2} & \dots & d_{n,m} \end{bmatrix}$$

Figura 1. Representação de um conjunto de dados.

Onde \vec{d}_n é um vetor de m coordenadas e n é o número de elementos do conjunto de dados. Cada vetor representa um dado desse conjunto e cada coordenada do vetor representa um atributo do dado. O conjunto de dados D reside no espaço \mathbb{R}^m , e este é referenciado como “espaço dos dados”, “espaço de entrada” ou “espaço vetorial” [10].

De um modo geral, os dados apresentam algum tipo de organização que pode ser representada através das partições do conjunto de dados. Segundo [10], estas partições podem ser chamadas de classes ou grupos de dados (*clusters*), a depender, respectivamente da existência ou não de algum tipo de informação discriminatória explícita sobre elas.

O termo classe deve ser utilizado quando existe informação sobre quantas e quais são as partições presentes no conjunto de dados e qual dado pertence a qual partição. Neste caso, a análise de dados é denominada classificação e propõe uma função que seja capaz de realizar o mapeamento entre os dados e suas

classes, além de mapear corretamente novos dados que possam vir a compor o conjunto de dados [10].

Já o termo grupo deve ser utilizado quando não existe informação sobre a organização dos dados. Neste caso, a análise dos dados é denominada agrupamento e tem por objetivo estudar as relações de similaridade entre os dados, determinando quais dados formam quais grupos [10].

A definição formal de agrupamento encontrada em [11] é: “Considerando o agrupamento de um conjunto de n objetos $D = \{\vec{d}_1, \vec{d}_2, \dots, \vec{d}_n\}$, estes devem ser agrupados em grupos distintos $G = \{G_1, G_2, \dots, G_c\}$, sendo c o número de grupos, de forma que as seguintes condições sejam atendidas:

1. $G_1 \cup G_2 \cup \dots \cup G_c = D$;
2. $G_i \neq \emptyset, \forall i, 1 \leq i \leq c$;
3. $G_i \cap G_j = \emptyset, \forall i \neq j, 1 \leq i, j \leq c$ ”.

Por estas condições, um objeto não pode pertencer a mais de um grupo (grupos distintos) e cada grupo tem que ter ao menos um objeto.

A tarefa de classificação se propõe a determinar uma função que seja capaz de realizar o mapeamento entre os dados e suas classes, bem como mapear corretamente novos dados que venham a compor o conjunto de dados [10].

B. Técnicas de Agrupamento

As técnicas de agrupamento têm como objetivo separar um conjunto de dados de modo que os elementos de um determinado grupo apresentem um grau de similaridade entre eles. Segundo [12], o uso destas técnicas é apropriado quando se conhece pouco ou nada sobre a estrutura do conjunto de dados a ser trabalhado.

Encontrar o melhor agrupamento para um conjunto de dados não é uma tarefa simples. Os autores de [13] afirmam que a maioria dos algoritmos requer valores para os parâmetros de entrada que são difíceis de determinar, além do fato de que os algoritmos são sensíveis a estes parâmetros, podendo produzir grupos muito diferentes de acordo com o ajuste destes valores.

O objetivo dos algoritmos que realizam o agrupamento de dados é encontrar duas ou mais coleções de dados que formam uma estrutura de grupo. Segundo [14], a ideia é que dados de um mesmo grupo tenham mais características em comum entre si do que com dados de outro grupo.

Entretanto, na prática, separar dados em grupos pode exigir a consideração de fatores de incerteza e imprecisão. [5] afirma que em muitos casos, assume-se que um elemento pertence a apenas um grupo, o que nem sempre ocorre em domínios reais, onde os elementos pertencem a mais de um grupo, com diferentes graus de pertinência.

Uma alternativa para o agrupamento de dados com sobreposição de grupos é a utilização de conjuntos *fuzzy*. Segundo os autores de [15], conjuntos *fuzzy* podem ser utilizados para representar os elementos do conjunto como um vetor de grau de pertinência, onde cada grau representa o grau de “posse” do atributo em questão, por parte dos elementos do grupo.

Com a utilização da abordagem *fuzzy*, o problema passa a ser de agrupamento *fuzzy*, cujo objetivo é obter partições *fuzzy* em um conjunto de dados D .

C. Medidas de Similaridade

Os algoritmos que executam tarefas de agrupamento de dados geralmente utilizam alguma medida de similaridade entre vetores em seu processo de execução. Segundo [10] estas medidas servem como direcionamento no processo de construção da superfície de decisão, que determinará qual ou quais dados pertencem a quais grupos.

A similaridade entre dois vetores corresponde a uma medida que compara a igualdade dos mesmos. De acordo com [16] esta medida também pode ser utilizada para medir a distinção entre os vetores (dissimilaridade).

Normalmente a similaridade é calculada com base em uma medida de distância espacial entre dois vetores. Existem diversas medidas de distâncias que podem ser utilizadas nos processos de agrupamento de dados: Distância Euclidiana, Distância de Hamming, Distância *City-Block*, Métrica de Tanimoto, Métrica de Mahalanobis, entre outros [10].

D. Fuzzy-c-Means

Proposto por Bezdek, Tsao e Pal [10], o algoritmo FCM é um dos mais atrativos para o agrupamento de dados, devido principalmente à sua simplicidade e eficiência.

O algoritmo FCM tem como objetivo encontrar grupos *fuzzy* para um conjunto de dados. Para alcançar este objetivo, o algoritmo precisa minimizar uma função que está relacionada à minimização das distâncias entre os dados e os centros dos grupos aos quais tais dados pertencem, com algum grau de pertinência. [17] Tal função é dada pela Eq. 1:

$$J_{FCM}(U_f, C) = \sum_{i=1}^c \sum_{j=1}^n (\mu_{ij})^m d(\vec{m}_i, \vec{x}_j)^2 \quad (1)$$

Onde n é o número de dados do conjunto de dados, c é o número de grupos a ser determinado pelo algoritmo, μ_{ij} indica a associação de um dado a um determinado grupo, m é um número real positivo utilizado como “parâmetro de fuzificação”, $d(\vec{m}_i, \vec{x}_j)^2$ é a distância entre o vetor de dados \vec{x}_j e o centroide representante da i -ésima partição *fuzzy* \vec{m}_i , U_f é uma matriz de pertinência *fuzzy*, C é uma matriz de centros de grupos e \vec{m}_i são centroides.

A atualização da matriz U_f acontece durante o processo de minimização da função objetivo (Eq. 1). A inicialização da matriz C é aleatória e sua atualização também ocorre durante o processo de minimização da função objetivo (Eq. 1).

O parâmetro de fuzificação m pondera a influência das relações de pertinência de um dado j a um grupo i em relação a todos os demais grupos, no cálculo das pertinências μ_{ij} . Quanto maior o valor de m , maior a atenuação das diferenças entre os graus de pertinência [10].

O algoritmo FCM utiliza como base do cálculo de similaridade entre vetores a Distância Euclidiana, que é dada pela Eq. 2:

$$d(\vec{v}_i, \vec{v}_j) = \left(\sum_{l=1}^p (v_{il} - v_{jl})^2 \right)^{\frac{1}{2}} \quad (2)$$

Onde p é a dimensão do espaço dos vetores e \vec{v}_i e \vec{v}_j são os vetores sobre os quais se deseja calcular a similaridade [10].

O processo de execução do FCM pode ser definido, segundo [18], como descrito na Fig. 2:

Fuzzy-c-Means
Determine o valor do parâmetro de “fuzificação” m ;
Determine a quantidade de partições <i>fuzzy</i> C ;
Determine um valor pequeno e positivo para o erro máximo, ϵ , permitido no processo;
Inicialize a matriz de centroides C aleatoriamente;
Inicialize o contador de iterações t como $t = 0$;
repita
$t++$;
Atualize U_f e C ;
até que $\ C^{(t)} - C^{(t-1)}\ < \epsilon$

Figura 2. Pseudocódigo do algoritmo de agrupamento Fuzzy-c-Means.

E. Subtractive Cluster

Proposto por [19], no algoritmo de classificação SC, cada ponto da base de dados é considerado um centroide de *cluster* em potencial. Adotando-se o critério de vizinhança, os pontos de dados que possuem muitos vizinhos em determinado raio de abrangência terão um potencial de alto valor. A forma para calcular esse potencial é vista na Eq. 3:

$$P_i = \sum_{j=1}^n e^{-\alpha \|x_i - x_j\|^2} \quad (3)$$

Onde x_i representa cada ponto de dado e α é mostrado na Eq. 4:

$$\alpha = \frac{4}{r_a^2} \quad (4)$$

r_a^2 é uma constante positiva que define o raio de vizinhança, e pontos de dados fora deste raio terão pouca influência no potencial.

Após ser calculado o potencial de cada ponto de dados, é selecionado o ponto de dados com maior potencial para ser o primeiro centroide de grupo. Sendo x_i^* o local do primeiro centroide de *cluster* e P_i^* seu valor potencial, o potencial de cada ponto x_i de dados é então revisado pela fórmula da Eq. 5:

$$P_i = P_i - P_i^* * e^{-\beta \|x_i - x_i^*\|^2} \quad (5)$$

Onde β é mostrado na Eq. 6:

$$\beta = \frac{4}{r_b^2} \quad (6)$$

Onde r_b é uma constante positiva e subtrai-se um valor de potencial sobre cada ponto de dados, como uma função de sua distância em relação ao primeiro centroide de *cluster*. Os pontos de dados próximos ao primeiro centroide de *cluster* terão seu potencial reduzido, tornando improvável serem eleitos como o

próximo centroide de *cluster*. Os pontos que estiverem no raio de abrangência do primeiro centroide pertencerão ao seu *cluster*.

Quando o potencial de todos os dados for revisado de acordo com a Eq.5, seleciona-se o ponto de dados com maior potencial remanescente como o segundo centroide de *cluster*. São realizadas sucessivas iterações, até que todos os pontos de dados estejam dentro do raio de algum centroide de *cluster*.

IV. BASES DE DADOS

Para realizar este trabalho, foram utilizadas, para criação dos modelos propostos, 3 bases de dados: a primeira contém informações sobre veículos novos, a segunda contém perfis de usuários e a terceira relaciona os perfis de usuários aos veículos que melhor se encaixam com o perfil informado. Estas bases serão detalhadas a seguir.

A. Base I – Veículos Novos

A primeira base utilizada foi a de automóveis, contendo 465 veículos novos a venda no Brasil, de dezessete fabricantes, incluindo as dez com maior volume de vendas anuais no país. Nesta base encontram-se todos os modelos disponíveis à venda por estas fabricantes. Além de suas respectivas marcas e modelos, também são registradas características de cada veículo, como número de portas, capacidade de passageiros, a potência (medida em cavalo-vapor), combustível utilizado, tipo de câmbio (transmissão), formato de carroceria e preço.

O formato de carroceria é um atributo fundamental na escolha do veículo. Ele define não apenas aspectos estéticos do desenho, como também aspectos funcionais, como capacidade de transporte de passageiros e de carga.

Há oito tipos de carrocerias na base de veículos: *hatch*, sedã, perua, picape, utilitário, monovolume, cupê e conversível.

Os veículos com carroceria *hatch* não possuem o porta-malas saliente, algo presente nos sedãs e diferenciando-os. As peruas são derivadas de sedãs, mas com maior capacidade de transporte no porta-malas.

As picapes possuem caçamba, têm maior aptidão para terrenos irregulares e algumas são voltadas para trabalho. Os utilitários possuem maior altura em relação ao solo. Os monovolumes são voltados para famílias que priorizam espaço interno. Esta carroceria possui teto mais alto.

Os cupês possuem um perfil mais esportivo, sempre com 2 portas. Os conversíveis são caracterizados por possuírem o teto removível, proporcionando um visual diferenciado em relação às outras carrocerias, também disponíveis apenas com 2 portas.

As carrocerias *hatch* e sedã são a maioria na base de dados, repercutindo a diversidade de opções disponíveis pelas fabricantes. Nestas carrocerias incluem-se desde *hatches* populares até sedãs luxuosos.

Veículos de nicho como cupês e conversíveis possuem maior faixa de preço e baixo volume de vendas, justificando a pouca variedade de opções presentes.

Peruas e monovolumes também possuem uma oferta reduzida, perdendo espaço no mercado frente aos utilitários. As picapes também vêm conquistando a preferência dos

consumidores, pois possuem uma gama de modelos diversificada no mercado.

A Tabela I mostra, de forma sucinta, todas as alternativas de carrocerias contidas na *Base I* com suas respectivas descrições, capacidade de transporte de passageiros e quantidade de carros correspondente a esta carroceria presentes na base.

TABELA I
INFORMAÇÕES ACERCA DOS ATRIBUTOS PRESENTES NA *BASE I*

Tipos de carroceria	Descrição	Capacidade de passageiros	Quantidade de carros cadastrados
HATCH	Não possuem porta-malas saliente	2, 4 e 5	158
SEDÃ	Possuem porta-malas saliente	2, 4 e 5	113
PERUA	Possuem maior capacidade de transporte no porta-malas	2, 4 e 5	12
PICAPE	Possuem caçamba, maior aptidão para terrenos irregulares e algumas são voltadas para trabalho.	2, 4, 5 e 6	72
UTILITÁRIO (SUV)	Possuem maior altura em relação ao solo.	2, 4 à 7	67
MONOVOLUME	Possuem grande espaço interno e teto mais alto.	2, 4 à 8	36
CUPÊ	Possuem um perfil mais esportivo.	2 e 4	3
CONVERSÍVEL	Possuem o teto removível.	2 e 4	4

A existência do número de portas nesta base é importante não apenas pelo fator visual da carroceria, mas também pela facilidade de acesso dos usuários ao automóvel. Os veículos normalmente possuem 2 ou 4 portas e na base de veículos deste trabalho, todos compartilham desta característica. Veículos com as carrocerias *hatch* e *picape* podem ter 2 ou 4 portas. Sedãs, peruas, utilitários e monovolumes possuem 4 portas, enquanto cupês e conversíveis possuem 2 portas.

O número de passageiros é relevante, pois informa o número da capacidade de transporte de pessoas no automóvel, incluindo o motorista. Este número pode variar de 2 pessoas, como picapes de cabine simples, até 8 pessoas, como monovolumes. A maioria dos veículos da base são capazes de transportar até 5 passageiros, desde *hatches* de 2 portas até peruas familiares. Apenas utilitários e monovolumes são capazes de transportar de 7 a 8 pessoas.

A capacidade de transporte de porta-malas, ou caçamba para a carroceria *picape*, é de grande importância para famílias que viajam com frequência, precisam de espaço para malas e outros

objetos, ou utilizam o veículo a trabalho no transporte de passageiros ou de cargas. As opções de capacidade de transporte do porta-malas são divididas em baixa, média ou alta.

A potência do motor é característica relevante para pessoas que fazem questão de um veículo mais ágil, ou que não se importam muito com desempenho, priorizando o consumo de combustível, por exemplo. A potência não foi exemplificada em números de cavalos-vapor do motor, pois muitos desconhecem estes números. As opções disponíveis são baixa, média ou alta potência. Geralmente veículos com alta potência tendem a ter preços mais elevados.

O combustível utilizado pelo veículo faz parte dos atributos, e apesar da maioria dos modelos da base de veículos adotarem combustível flex (gasolina e álcool), há quem prefira motores exclusivamente a gasolina ou diesel, este último conhecido por economizar custos após um período de quilometragem percorrida com o automóvel. Somente utilitários e picapes de maior porte aceitam combustível diesel.

O tipo de câmbio está presente como atributo porque há pessoas que preferem o câmbio manual ao automático, e vice-versa. O câmbio automático prioriza o conforto, dispensando o motorista de realizar as trocas de marchas manualmente, porém eleva o preço final do automóvel. O câmbio manual possui preço de aquisição mais barato, não prioriza o conforto, mas remete a um estilo de condução do veículo mais esportivo e prazeroso para algumas pessoas.

A faixa de preço limita o número e variedade de opções do usuário, de acordo com a quantia que este deseja investir. Há cinco faixas de preço neste trabalho, são elas: 21 a 36 mil reais, a partir de 36 até 50 mil reais, a partir de 50 até 70 mil reais, a partir de 70 até 100 mil reais, e acima de 100 mil reais. É observável que veículos mais velozes, luxuosos e equipados estão em faixas de preços mais elevadas.

B. Base II – Perfis de Consumo

Para a criação da segunda base de dados, composta por perfis de consumo de usuários, foi utilizado um formulário eletrônico de pesquisa de opinião, disponível em: <http://goo.gl/forms/XTLAuoulpf>. Através dele, os usuários preencheram as informações desejadas sobre as características do automóvel, explicadas anteriormente, e também informações pessoais sobre seu perfil.

Os atributos sexo e idade foram escolhidos para ajudar na elaboração do perfil do usuário. A idade foi dividida em faixas etárias que variam da seguinte forma: até 20 anos, entre 21 e 30 anos, entre 31 e 40 anos, entre 41 e 59 anos, e acima de 60 anos. Pessoas mais jovens tendem a escolher veículos mais baratos, pois geralmente estão em período de estudos ou no começo de suas carreiras. Pessoas com maior idade estão inseridas na faixa de compradores de veículos mais caros, ou conforme será descrito a seguir, possuem família e necessitam de veículos mais específicos.

O estado civil aliado ao número de filhos também foram escolhidos como atributos, visto que a quantidade de pessoas que moram com o consumidor do veículo influencia nas decisões sobre características do veículo, como número de portas, capacidade de passageiros, porta-malas e tipo de

carroceria. Famílias com muitos filhos e parentes morando juntos tendem a escolher veículos mais espaçosos como peruas, utilitários ou monovolumes.

Também foi selecionada como atributo, a escolha entre utilizar o veículo para lazer (passeios ou deslocar ao trabalho) ou para trabalho (transporte de carga ou passageiros, realização de fretes). Em geral, pessoas que utilizam o veículo para trabalho priorizam características “racionais” como consumo de combustível, capacidade de transporte de passageiros e carga, ou custo de aquisição e manutenção. Em casos onde o automóvel é utilizado como ferramenta de trabalho, o menor custo de aquisição e manutenção reflete em maior lucratividade.

Adicionalmente foi questionado qual é o fator mais importante a ser considerado no momento da compra do veículo. As opções são: espaço interno, consumo, *design*, manutenção barata, segurança, conforto e desempenho. As respostas refletem as preocupações que o usuário tem durante a escolha, afinal, escolher uma das opções indica se o usuário prioriza equipamentos de comodidade ou segurança, peças de reposição baratas, motores econômicos ou muito potentes, um desenho atraente da carroceria, ou amplo espaço interno.

A Tabela II mostra todos os atributos que compõem a *Base II* e suas respectivas alternativas possíveis de resposta:

TABELA II
INFORMAÇÕES ACERCA DOS ATRIBUTOS PRESENTES NA *BASE II*

Atributos	Alternativas possíveis
Sexo	Masculino
	Feminino
	Outro
Idade	Até 20
	De 21 a 30
	De 31 a 40
	De 41 a 59
	Acima de 60
Estado Civil	Solteiro (a)
	Casado (a)
	Divorciado (a)
	Viúvo (a)
Número de Filhos	Nenhum
	1
	2
	3
	4 ou +
Mora com os Filhos	Sim
	Não
Utilização para Lazer ou Trabalho	Lazer (Passeio, ir ao trabalho)
	Trabalho (Transporte de carga/passageiros e serviço para empresas)
Fator Determinante	Desempenho
	Conforto
	Segurança
	Manutenção barata
	Consumo
	Espaço interno
	<i>Design</i>

C. Base III – Relacionamento de Perfis e Veículos

A terceira base relaciona para cada perfil de consumo da *Base II*, um máximo de quatro automóveis, dentre os

cadastrados na *Base I*, que satisfaçam as necessidades e desejos de cada perfil de consumo.

A oferta e variedade de automóveis “populares” são maiores do que a de esportivos luxuosos, e as fabricantes proporcionam uma diversidade de modelos relativa à demanda do mercado. A fim de manter esta proporção na base criada, considerando duas pessoas com perfis semelhantes, mas com diversas opções de modelos que satisfaçam seu perfil, foram escolhidos modelos que satisfaçam ambos, porém não exatamente iguais para as duas pessoas.

Por exemplo, supondo dois perfis de consumo semelhantes, que dentre os 465 veículos cadastrados na Base I, os carros “populares” A, B, C, D, E e F sejam os mais adequados. Considerando estes dois perfis, para o primeiro poderiam ser associados os veículos A, F, E e B, enquanto para o segundo poderiam ser associados os veículos B, C, E e D.

Esta estratégia de associação garante a representação de diversas opções de modelos e evita exclusão ou “favorecimento” de marcas ou modelos.

Em alguns casos de perfis específicos onde há pouca variedade de modelos indicados, um número menor de sugestões foi atribuído.

V. METODOLOGIA

A primeira etapa de desenvolvimento deste trabalho consistiu na obtenção do conjunto de perfis de usuários. Para tal, foi necessário determinar, dentre os 14 atributos da *Base II*, quais representariam o perfil de um consumidor de automóveis novos.

Para que o algoritmo FCM realize o agrupamento, além dos perfis de usuários, é necessário saber em quantos grupos o conjunto de dados deverá ser dividido. A fim de obter esta informação, o algoritmo SC foi aplicado no conjunto de perfis de usuários e este retornou a quantidade de grupos “ideal” para este conjunto de dados.

Tendo como entrada o conjunto de perfis de usuários e o número de grupos, o algoritmo FCM realizou o agrupamento obtendo duas saídas: a matriz de pertinência e o vetor de centroides. Estas duas informações, combinadas, são utilizadas para encontrar qual grupo melhor representa um novo perfil de consumidor.

Com o conhecimento do grupo que melhor se adequa ao novo perfil de consumidor, os carros vinculados aos demais perfis deste grupo também estarão associados ao novo perfil; tendo então, possíveis carros que são compatíveis ao que o usuário precisa. Dentre todos os carros que atenderiam aos desejos e necessidades do usuário, serão sugeridos um máximo de três carros.

A Fig. 3 mostra a metodologia utilizada na execução do trabalho.

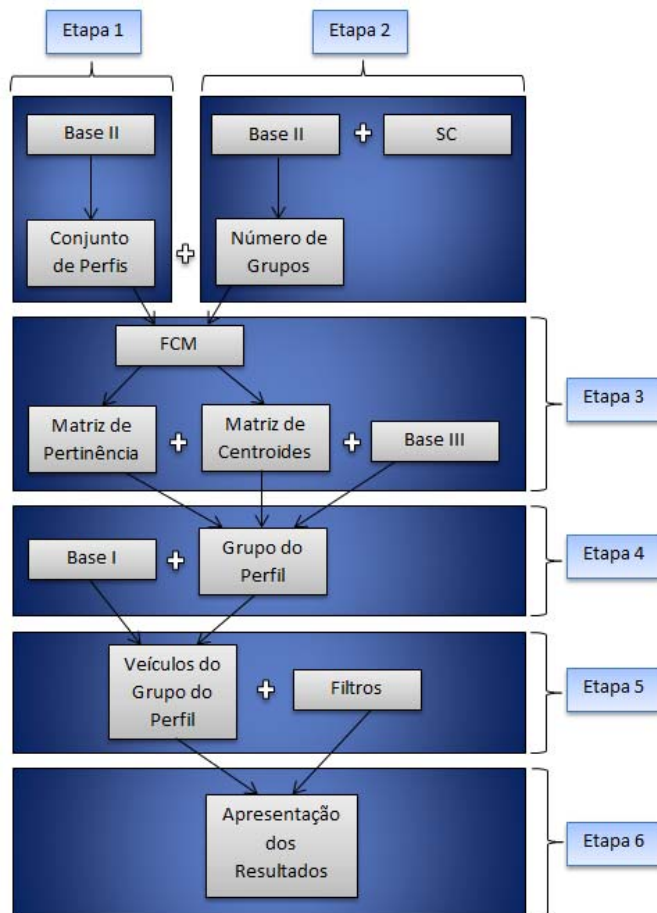


Figura 3. Fluxograma da metodologia utilizada.

Para melhor entendimento do desenvolvimento do trabalho, a sua metodologia será descrita em seis etapas, sendo elas:

1. Obtenção do conjunto de perfis de usuários;
2. Obtenção do número de grupos;
3. Realização do agrupamento;
4. Obtenção do grupo do usuário;
5. Obtenção dos carros do grupo do usuário;
6. Apresentação dos resultados.

A. Obtenção do conjunto de perfis de usuário

A primeira etapa de desenvolvimento deste trabalho foi determinar entre os atributos da *Base II*, quais representariam o perfil de um consumidor de automóveis novos.

Primeiramente, todos os 14 atributos que compõe a *Base II* foram utilizados para formar o perfil de um usuário, porém, notou-se que os grupos formados no agrupamento apresentavam alta similaridade, não diferenciando os perfis de consumo de maneira efetiva.

A fim corrigir este problema foram escolhidos atributos, que combinados, fizeram com que os grupos retornados pelo FCM apresentassem uma maior dissimilaridade entre si, representando uma maior variedade de perfis em grupos distintos.

Dentre todos os atributos, os selecionados foram: o número de filhos da pessoa e se estes moram com ela, número de portas do veículo, as capacidades de passageiros e porta-malas,

potência do motor, faixa de preço de aquisição do automóvel e seu tipo de carroceria.

A Tabela III mostra todos os atributos que foram escolhidos para representar o perfil de um usuário.

TABELA III
ATRIBUTOS ESCOLHIDOS PARA REPRESENTAR O PERFIL DE UM USUÁRIO

Atributos
Número de filhos
Mora com os filhos
Número de portas
Capacidade de passageiros
Potência do motor
Capacidade do porta-malas
Faixa de preço
Tipo de carroceria

Na visão de um especialista na área, o número de filhos e se estes moram com os usuários são atributos que podem influenciar na escolha da capacidade de número de passageiros, capacidade do porta-malas, carroceria e número de portas, pois estes atributos influenciam no acesso, comodidade e conforto dos usuários.

A potência do motor pondera a predileção por desempenho maior ou consumo de combustível menor. A capacidade do porta-malas é relevante quando o usuário necessita de espaço para guardar bagagens e malas, que podem crescer de acordo com o número de integrantes da família.

O tipo de carroceria caracteriza o formato e perfil do modelo, as opções podem atrair por aspectos funcionais como melhor capacidade de transporte de passageiros ou carga, e aspectos emocionais como esportividade, elegância ou imponência da carroceria.

A faixa de preço limita quanto o usuário pretende investir. Pessoas com maior poder aquisitivo tendem a optar por modelos mais potentes, luxuosos, equipados e confortáveis. Faixas de preço abaixo de R\$50 mil reais impedem a aquisição de modelos cupês, conversíveis ou que utilizem combustível diesel, por exemplo.

Estes atributos considerados relevantes na visão do especialista, combinados, representam um perfil de usuário. A junção de todos os perfis de usuário compõe o conjunto de perfis de usuários.

B. Obtenção do número de grupos

Além do conjunto de perfis de usuários, outra entrada do algoritmo FCM, como mostrado na Fig. 2, é a quantidade de partições *fuzzy*, ou seja, o número de grupos que o algoritmo irá segregar os dados.

Saber o número de grupos exato que garantirá um resultado satisfatório no agrupamento não é uma tarefa simples, pois é desejável que os grupos resultantes tenham homogeneidade entre seus elementos, e heterogeneidade entre os grupos.

Nesta segunda etapa, a fim de se obter a quantidade de partições *fuzzy* que será utilizada, foi utilizado o algoritmo SC, que retornou o número de grupos “ideal”, utilizando como entrada o conjunto de perfis de usuários.

C. Realização do Agrupamento

O algoritmo de agrupamento FCM necessita de entradas numéricas para seu funcionamento. Para atender tal exigência, cada atributo relevante que compõe um perfil de usuário (Tabela III) foi mapeado para números inteiros de acordo com as alternativas das respostas. A Tabela IV mostra este mapeamento.

Supondo um perfil de usuário que possua 1 filho e este não more com ele, queira um carro com 4 portas, capacidade de porta malas pequena, motor de potência média, com capacidade para 5 passageiros, faixa de preço de R\$ 36.001 a 50.000 e que seja *hatch*. O mapeamento para este perfil seria o seguinte: 1 1 4 0 1 5 1 0. Este mapeamento para números inteiros é feito em todo o conjunto de perfis de usuários.

Com o número de grupos definido e o conjunto de perfis mapeado para dados numéricos, inicia-se a terceira etapa, na qual o FCM realiza o agrupamento da base de dados de perfis de usuários. O resultado da execução deste algoritmo é uma matriz contendo os centroides de cada grupo e uma matriz com valores finais da partição fuzzy, também conhecida como matriz de pertinência.

A matriz de centroides contém em cada linha, valores que representa os perfis de um determinado grupo. A matriz de pertinência contém, para cada perfil do conjunto de perfis de usuários, uma porcentagem indicando a similaridade destas com cada um dos centroides dos grupos formados.

D. Obtenção do Grupo do Usuário

Depois de obtidas as matrizes dos centroides e de pertinência, foi necessário encontrar o grupo ao qual o novo usuário pertence. A Tabela V mostra uma matriz de pertinência contendo cinco perfis de usuários (representando o conjunto de perfis de usuários) e três grupos (representando os grupos formados pelo algoritmo FCM).

TABELA V
EXEMPLO DE UMA MATRIZ DE PERTINÊNCIA PARA CINCO PERFIS DE USUÁRIOS E TRÊS GRUPOS

Centroides	Perfil 1	Perfil 2	Perfil 3	Perfil 4	Perfil 5
Centro 1	0.24	0.40	0.34	0.08	0.55
Centro 2	0.48	0.50	0.30	0.90	0.20
Centro 3	0.28	0.10	0.36	0.02	0.25

Observando o perfil de usuário 1, percebe-se que este possui 0.24 de semelhança com o centro 1, 0.48 com centro 2 e 0.28 com o centro 3. Como este usuário possui maior semelhança com o centro 2 (48%), ele passa a pertencer ao grupo 2. Fazendo a mesma análise para os outros usuários, o grupo 1 receberá somente o usuário 5. O grupo 2, que já contém o usuário 1, receberá os usuários 2 e 4 e o grupo 3 receberá o usuário 3.

Seguindo este mesmo raciocínio é que um novo usuário será associado a um dos 15 grupos retornados pelo algoritmo FCM.

E. Obtenção dos Carros do Grupo do Usuário

Na quinta etapa, com o conhecimento do grupo ao qual o usuário pertence, os carros vinculados aos integrantes deste grupo agora também estarão associados ao novo usuário. Como o novo usuário e os integrantes do grupo ao qual ele pertence possuem perfis semelhantes, espera-se que os veículos previamente associados aos integrantes do grupo também o satisfaçam.

É necessário ressaltar, que as pessoas foram agrupadas com perfis de consumo parecidos, não necessariamente idênticos. Em um mesmo grupo, podemos ter usuários com mais de dois filhos que optaram por uma perua, enquanto outros, por um monovolume. Esta divergência de respostas por pessoas de mesmo perfil pode “frustrar” o usuário da aplicação, que cobiça adquirir uma perua, mas acaba recebendo sugestões com outro tipo de carroceria pela aplicação.

F. Apresentação dos resultados

Nesta sexta etapa, antes de apresentar as três sugestões de carros ao usuário, com o intuito de melhorar a fidelidade entre as opções preenchidas por este e os veículos retornados, foram realizados filtros.

O primeiro filtro utilizado é o de retornar somente automóveis da carroceria escolhida pelo usuário, excluindo os que destoam deste aspecto visual.

Se mesmo após a realização do primeiro filtro, ainda existirem mais de três opções de veículos, outros filtros são realizados na seguinte ordem: por tipo de câmbio, por número de portas e por preço. Estes atributos tiveram sua importância explicada anteriormente e foram suficientes para sugerir veículos condizentes com o perfil consumidor, além de diminuir a quantidade de sugestões.

Ainda sim, se mesmo após a realização dos quatro filtros (por carroceria, câmbio, número de portas e preço), existirem mais de três veículos adequados, é realizado um sorteio entre estes veículos, e as três opções sorteadas são apresentadas ao usuário, concluindo a última etapa.

Este sorteio possibilita que um mesmo usuário, ao executar a aplicação mais de uma vez, possa receber resultados diferentes.

Caso existam menos de três opções que se enquadram no perfil do usuário, estas são apresentadas; e se não existir um veículo com as especificações solicitadas, o usuário é informado.

Para melhor entendimento das etapas descritas, o Quadro I apresenta uma descrição detalhada do algoritmo desenvolvido.

QUADRO I. DETALHAMENTO DO ALGORITMO PROPOSTO.

```

1. Solicita (sexo, idade, estado_civil, n_filhos, mora_filhos, lazer_ou_trabalho, fator_determinante);
2. Solicita
(n_portas, capacidade_porta_malas, capacidade_passageiros, potencia_motor, combustivel, cambio, faixa_preço, car
roceria);
3. perfil_de_usuario <- Cria_Perfil
(n_filhos, mora_filhos, n_portas, capacidade_porta_malas, capacidade_passageiros, potência_motor, faixa_preço, c
arroceria);
4. numero_grupos <- Algoritmo_SC (perfil_de_usuario);
5. perfil_numerico <- Mapeamento (perfil_de_usuario);
6. centroides, matriz_pertinencia <- Algoritmo_FCM (numero_grupos, perfil_numerico);
7. grupo_usuario <- Encontra_Grupo_Usuario (centroides, matriz_pertinencia, perfil_numerico);
8. carros_grupo_usuario <- Recupera_Carros (grupo_usuario);
9. n_carros_grupo_usuario <- Conta_Carros (carros_grupo_usuario);
10. Se (n_carros_grupo_usuario = 0) entao
11. Mostra ("Não existem carros que se enquadram às informações informadas!");
12. Senao
13. Se (n_carros_grupo_usuario <= 3) entao
14. Mostra (carros_grupo_usuario);
15. Senao
16. carros_filtrados_carroceria <- Filtrar_Carroceria (carros_grupo_usuario);
17. carros_filtrados_cambio <- Filtrar_Cambio (carros_filtrados_carroceria);
18. carros_filtrados_n_portas <- Filtrar_N_Portas (carros_filtrados_cambio);
19. carros_filtrados_finais <- Filtrar_Preço (carros_filtrados_n_portas);
20. n_carros_filtrados_finais <- Conta_Carros (carros_filtrados_finais);
20. Se (n_carros_filtrados_finais > 3) entao
21. carros_finais <- Sorteia3 (carros_filtrados_finais);
22. Mostra (carros_finais);
23. Senao
24. Mostra (carros_filtrados_finais);

```

VI. ESTUDO DE CASO

A ideia básica do agrupamento consiste em colocar em um mesmo grupo objetos que sejam similares de acordo com um critério pré-determinado.

Neste contexto deste trabalho, o agrupamento do conjunto de perfis de consumo é relevante no sentido de determinar quais

são os usuários que apresentam um perfil semelhante, considerando como critério as motivações na hora da compra de um carro novo.

Para um melhor entendimento dos 15 grupos - obtidos através do agrupamento da *Base II* - foi realizada uma análise dos perfis de usuários de cada grupo:

TABELA IV

MAPEAMENTO DOS ATRIBUTOS RELEVANTES: POTÊNCIA DO MOTOR, CAPACIDADE DE PASSAGEIROS, FAIXA DE PREÇO E TIPO DE CARROCERIA, NÚMERO DE FILHOS E SE MORA COM OS FILHOS, NÚMERO DE PORTAS E CAPACIDADE DO PORTA-MALAS

Mapeamento Das Entradas	Potência Do Motor	Capacidade De Passageiros	Faixa De Preço	Carroceria	Número De Filhos	Mora com os Filhos	Número de Portas	Capacidade do Porta Malas
0	Pouco Potente	-	21.000 a 36.000	Hatch	Nenhum	Sim	-	Pequeno
1	Potência Média	-	36.001 a 50.000	Sedan	1	Não	-	Médio
2	Muito Potente	2	50.001 a 70.000	Perua	2	-	2	Grande
3	-	-	70.001 a 100.000	Picape	3	-	-	-
4	-	4	Acima de R\$100.000	Utilitário (SUV)	4 ou +	-	4	-
5	-	5	-	Monovolume	-	-	-	-
6	-	6	-	Cupê	-	-	-	-
7	-	7	-	Conversível	-	-	-	-
8	-	8 ou +	-	-	-	-	-	-

A. Análise dos grupos formados

Os atributos dos perfis analisados foram: número de filhos e se moram com os pais, número de portas do veículo, capacidade de passageiros, potência do motor, capacidade de porta malas, faixa de preço e tipo de carroceria.

Analisando o número de filhos e se estes moram com os pais, o único grupo que possui exclusivamente usuários que não

possuem filhos é o grupo 6. Os grupos 3, 8 e 13 possuem usuários que não tem filhos ou possuem somente um filho. Já os grupos 2 e 9 representam os usuários que possuem de 3 a 4 filhos, porém a maioria dos filhos dos integrantes do grupo 2 não moram com eles, enquanto no grupo 9, sim. Os grupos 14 e 15 possuem integrantes com no máximo 2 filhos e os grupos

1, 4, 5, 7, 10, 11 e 12 são caracterizados por representarem usuários com no máximo 3 filhos.

Com relação ao número de portas do veículo, 7 dos 15 grupos apresentaram usuários com preferência de carros com 4 portas, exclusivamente. Os 8 grupos restantes possuem usuário com preferência tanto de 2 quanto 4 portas. O grupo que possui a maior porcentagem (82,35%) de usuários que preferem um veículo com 2 portas é o 5, já o grupo que possui a menor porcentagem é o 4, com apenas 0,97%.

Quanto a capacidade de passageiros, os grupos 1, 3, 4, 5, 14 e 15 representam usuários que escolheram carros que transportam até 5 passageiros. O grupo 5 só possui integrantes que escolheram carros com capacidade para 6 passageiros, enquanto os grupos 2, 7, 10, 11, 12 e 13 possuem capacidade de transporte de 4 a 7 passageiros. Já o grupo 9 é o único com integrantes que necessitam transportar 8 ou mais pessoas.

A maioria dos grupos apresentam representantes das três opções de potência: baixa, média e alta, no entanto os grupos 1 e 5 apresentam integrantes que tem preferência por potências menores (de baixa e média) e os integrantes dos grupos 6, 8, 9, 10, 12 e 15 preferem potências mais altas (de média a alta).

Todos os integrantes dos grupos 6 e 8 tem preferência por porta-malas com capacidade de carga média. Os grupos 9, 10 e 11 representam pessoas que preferem porta-malas de porte médio a grande, enquanto o grupo 5 possui integrantes que têm preferência por porta-malas pequenos e médios. O grupo que possui a maior porcentagem de integrantes que prefere porta-malas grande é o 10, com 82,75% e a maior porcentagem de pessoas que preferem por porta-malas pequenos é a do grupo 1, com 56,69%.

As faixas de preço apresentaram grande diversidade entre os grupos. Os grupos 1 e 13 possuem integrantes que estão dispostos a gastar de 21 a 50 mil reais na compra de um carro zero. Os grupos 2, 7 e 11 possuem pessoas que não gastariam mais de 100 mil reais na compra de um carro. Já as pessoas que compõe o grupo 6 se enquadram na faixa de preço de 21 a 36 mil reais, enquanto os integrantes do grupo 8 na faixa de 36 a 50 mil.

Considerando as opções de carroceria, 9 dos 15 grupos são compostos por *hatches*, sendo o grupo 13 composto somente por usuários que preferem esta carroceria. O grupo 6 representa usuários que têm preferência por sedãs. Peruas, picapes e utilitários são representados pelos grupos 9, 10 e 11, cupês pelos grupos 11, 14 e 15 e os três grupos que contém conversíveis são o 9, 14 e 15.

O algoritmo de agrupamento FCM foi testado com diferentes perfis de consumo a fim de verificar a eficiência do mesmo e apresentou, de modo geral, dois tipos de resultados: quando se tem certeza da pertinência de um usuário em um único grupo e a possibilidade de pertinência em mais de um grupo por um mesmo usuário. Os resultados apresentados foram os seguintes:

B. Resultado com pertinência em um único grupo

No estudo de caso a seguir, foram utilizadas as seguintes informações: homem divorciado com 60 anos ou mais, com 2 filhos que moram com ele, o veículo seria utilizado para lazer,

priorizando desempenho na hora da compra. O veículo desejado deve ter 4 portas, capaz de transportar 5 passageiros, além de possuir porta-malas grande. Foi selecionado motor de alta potência, combustível apenas gasolina, e sedã de câmbio automático com faixa de preço superior a 100 mil reais.

Neste caso, o usuário é mais parecido com os integrantes do grupo 12. Na Fig. 5 é exibida a porcentagem de pertinência por grupo desta simulação:

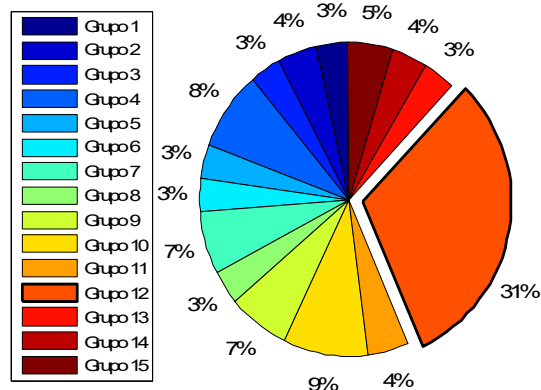


Figura 5. Gráfico do estudo de caso 1 com pertinência em um único grupo.

Para este perfil de consumo, há mais de três modelos de veículos compatíveis. A aplicação retornou dentre estas possibilidades, três automóveis condizentes com cada atributo informado pelo usuário.

Para este estudo, as opções sugeridas fazem parte de um mercado de nicho, com volume de vendas baixo e menor variedade de opções disponíveis pelas fabricantes. Este fato se dá por conta da faixa de preço escolhida (acima de 100 mil reais).

C. Resultado com pertinência em mais de um grupo

Nesta simulação, a pessoa possui as seguintes características: homem solteiro com idade entre 21 e 30 anos, não possui filhos, o veículo seria utilizado para lazer, e o consumo é determinante no momento da compra. O automóvel desejado possui 4 portas, porta-malas pequeno, capacidade de transportar 4 passageiros, motor que utiliza combustível flex e pouco potente, sendo um *hatch* de câmbio manual custando entre 21 e 36 mil reais.

Para este caso, o grupo 1 foi escolhido, como mostra a Fig. 6. Este perfil apresentou a mesma porcentagem de semelhança com seis grupos, porém os atributos que foram determinantes para que o grupo 1 fosse escolhido foram: o número de filhos e o número de portas. A maioria dos integrantes deste grupo possuem preferência por carros 4 portas e não possuem filhos, o que confere com o perfil do usuário informado.

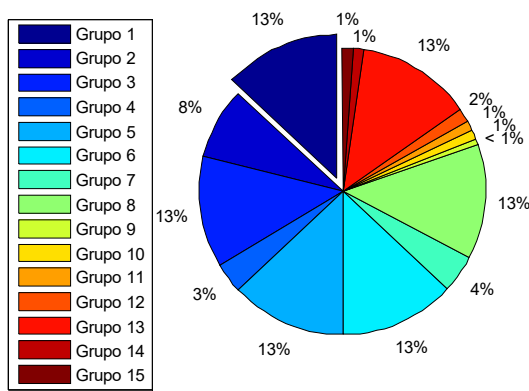


Figura 6. Gráfico do estudo de caso 2 com pertinência em mais de um grupo.

Neste segundo estudo de caso, a aplicação apresentou três opções de modelos de diferentes marcas que respeitam os atributos informados pelo usuário.

A categoria de veículos que foi sugerida para este perfil possui grande oferta de modelos pelas fabricantes, por esta é composta por automóveis mais acessíveis para o mercado consumidor e com maior volume de vendas.

VII. CONCLUSÃO

O contexto escolhido, consumo de veículos novos, fornece um panorama complexo o suficiente para justificar pesquisas que possam balizar sugestões de automóveis, respeitando as preferências e características dos consumidores.

Este trabalho propôs uma metodologia para identificar, entre grupos de perfis de usuários, a qual grupo um novo perfil deverá ser associado e tendo esta informação e os automóveis que se adequam a este grupo, sugerir veículos que satisfaçam os desejos e necessidades deste novo perfil.

Em casos onde o perfil recebe muitas sugestões de veículos, a aplicação de filtros de carroceria, tipo de câmbio, número de portas e preço permitiram afunilar o catálogo de opções, aumentando a fidelidade entre o que foi informado pela pessoa e as indicações de automóveis retornadas.

Nos estudos de caso realizados, alguns perfis específicos não foram tão bem distinguidos entre os grupos formados. Diferenças muito pequenas de pertinência (menos de 1%) ocorreram entre grupos mais parecidos, enquanto outros apresentaram maior semelhança (mais de 30%) com o centroide de um dos grupos.

Foi observado em ambas as simulações do estudo de caso, a aplicação sugeriu automóveis condizentes com as características do perfil de consumo, mostrando que este trabalho possibilitou nortear opções de veículos para perfis de consumo variados, cumprindo com o objetivo de facilitar o processo decisório de compra de um novo automóvel.

AGRADECIMENTOS

Os autores gostariam de agradecer aos integrantes do Laboratório de Inteligência Computacional (LInC) e a Universidade Federal de Alfenas (UNIFAL-MG) pelo apoio e

disponibilização da infraestrutura para o desenvolvimento da pesquisa.

REFERÊNCIAS

- [1] J. W. Newman and R. Staelin, *Information Sources of Durable Goods*, Journal of Advertising Research, New York, v. 13, p. 19-29, Apr. 1973.
- [2] ANFAVEA ANUARIO 2014, pg 15. [acesso 2014-09-08]. Disponível em: <<http://www.virapagina.com.br/anfafea2014/>>
- [3] LIU, W. Y., Xiao, C. J., Wang, B. W., Shi, Y., & Fang, S. F. (2003, Novembro). Study on combining subtractive clustering with fuzzy c-means clustering. In Machine Learning and Cybernetics, International Conference on (Vol. 5, pp. 2659-2662). IEEE. 2003.
- [4] PONCHIO, Mateus Canniatti; ARANHA FILHO, Francisco José Epósito; SAMARTINI, André Luiz Silva. Uso de fontes de recomendação fortes e fracas na escolha de veículo. 2002.
- [5] SALGADO, Ricardo M., OHISHI, Takaaki., BALLINI, Rosângela. Clustering Bus Load Curves. PSCE – Power System Conference Exposition/IEEE. IEEE Power Engineering Society. New York City, USA. 2004.
- [6] SASTRE, PT de N.; SERRALVO, F.; MORAS, A. C. A Influência dos Grupos de Referência no Processo Decisório do Consumidor: um estudo no segmento de veículos utilitários esportivos. Dissertação de Mestrado. Pontifícia Universidade Católica de São Paulo. São Paulo. 2009.
- [7] AZEVEDO, Bárbara Brzezinski. Agrupamento de trabalhadores com perfis semelhantes de aprendizado utilizando técnicas multivariadas. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul. Escola de Engenharia. Porto Alegre. 2013.
- [8] S de Oliveira Sampaio, D., Viscardi, A. W., Ornelas, R., do Nascimento, A. F., de Castro, C. V., Afonso, D. F., ... & da Silva, R. F. T. Um Estudo Comparativo sobre o Comportamento do Consumidor de Automóveis Novos. VII Seminário em Administração. FEA/USP. São Paulo. 2004.
- [9] de Sena Nunes, N., Pinto, L. A., Murakami, L. C., & Pimenta, M. L. O carro e os jovens: um estudo de valores através do método da Cadeia de Meios e Fins. XV Congresso Latino-Iberoamericana de Gestão de tecnologia. Portugal. 2013.
- [10] S ROCHA, Thiago, PERES, Sarajane M., BÍSCARO, Helton H., MADEO, Renata Cristina B., BOSCARIOLI, Clodis. Tutorial sobre Fuzzy-c-Means e Fuzzy Learning Vector Quantization: Abordagens Híbridas para Tarefas de Agrupamento e Classificação. Revista de Informática Teórica e Aplicada, Rio Grande do Sul, v.19, n. 1, p. 120-163, 2012.
- [11] HRUSCHKA, E. R. e EBECKEN, N. F. A Genetic algorithm for cluster analysis, IEEE Transactions on Evolutionary Computation, 2001. *Revista Eletrônica de Iniciação Científica*, v. 1, p. 1-10. 2006.
- [12] BACKER, E. Computer Assisted Reasoning in Cluster Analysis. Prentice Hall, New York, 1995.
- [13] ANKERST, M., BREUNIG, M. e KRIEGLER, H. Optics: Ordering Points to identify the clustering structure, Proceedings of the ACM SIGMOD Conference on Management of Data pp.49-60, 1999.
- [14] OLIVEIRA, Tatyana Bitencourt Soares de. *Clusterização de dados utilizando técnicas de redes complexas e computação bioinspirada* [online]. São Carlos : Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2008. Dissertação de Mestrado em Ciências de Computação e Matemática Computacional. [acesso 2014-10-09]. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-01042008-142253/>>.
- [15] KLIR, G. J. e YUAN, B. Fuzzy Sets and Fuzzy Logic Theory and Applications, Prentice Hall, 1995.
- [16] EVERITT, B. S. E RABE-HESKETH, S. The Analysis of Proximity Data. Hodder Arnold Publishers, Londres, 1997.
- [17] XU, R. E WUNSCH, D. II survey of clustering algorithms. IEEE Transactions on Neural Networks 16, 3, 645–678, maio de 2005.
- [18] BEZDEK, J. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
- [19] CHIU, Stephen L. Fuzzy model identification based on cluster estimation. Journal of intelligent and Fuzzy systems, v. 2, n. 3, p. 267-278, 1994.



Fernando Augusto Scaranelli Vince, possui graduação em Ciência da Computação pela Universidade Federal de Alfenas (2014). Tem experiência na área de Ciência da Computação.



Leandra de Carvalho Nogueira, é graduada em Ciência da Computação pela Universidade Federal de Alfenas – UNIFAL-MG. Atualmente é integrante do Laboratório de Inteligência Computacional – LinC.



Ricardo Menezes Salgado, é bacharel em Matemática pela Universidade Federal de Viçosa, mestre e doutor em Engenharia Elétrica pela Universidade Estadual de Campinas. Atualmente é professor do Núcleo de Ciência da Computação da Universidade Federal de Alfenas – UNIFAL - MG. Atua na área inteligência computacional, sendo consultor de diversos projetos em grandes empresas. Possui larga experiência na área de sistemas inteligentes, análise de dados e descoberta de conhecimento.