

APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS

Julio C. Uzinski
Caio C. E. de Abreu
Bruno R. de **Oliveira**

organizadores



Pantanal Editora

2020

Julio Cezar Uzinski
Caio Cesar Enside de Abreu
Bruno Rodrigues de Oliveira
(Organizadores)

**Aplicações de
Inteligência Artificial e
Ciência de Dados**



Pantanal Editora

2020

Copyright© Pantanal Editora
Copyright do Texto© 2020 Os Autores
Copyright da Edição© 2020 Pantanal Editora
Editor Chefe: Prof. Dr. Alan Mario Zuffo
Editores Executivos: Prof. Dr. Jorge González Aguilera
Prof. Dr. Bruno Rodrigues de Oliveira

Diagramação: A editora
Edição de Arte: A editora
Revisão: Os autor(es), organizador(es) e a editora

Conselho Editorial

- Prof. Dr. Adaylson Wagner Sousa de Vasconcelos – OAB/PB
- Profa. Msc. Adriana Flávia Neu – Mun. Faxinal Soturno e Tupanciretã
- Profa. Dra. Albys Ferrer Dubois – UO (Cuba)
- Prof. Dr. Antonio Gasparetto Júnior – IF SUDESTE MG
- Profa. Msc. Aris Verdecia Peña – Facultad de Medicina (Cuba)
- Profa. Arisleidis Chapman Verdecia – ISCM (Cuba)
- Prof. Dr. Bruno Gomes de Araújo - UEA
- Prof. Dr. Caio Cesar Enside de Abreu – UNEMAT
- Prof. Dr. Carlos Nick – UFV
- Prof. Dr. Claudio Silveira Maia – AJES
- Prof. Dr. Cleberton Correia Santos – UFGD
- Prof. Dr. Cristiano Pereira da Silva – UEMS
- Profa. Ma. Dayse Rodrigues dos Santos – IFPA
- Prof. Msc. David Chacon Alvarez – UNICENTRO
- Prof. Dr. Denis Silva Nogueira – IFMT
- Profa. Dra. Denise Silva Nogueira – UFMG
- Profa. Dra. Dennyura Oliveira Galvão – URCA
- Prof. Dr. Elias Rocha Gonçalves – ISEPAM-FAETEC
- Prof. Me. Ernane Rosa Martins – IFG
- Prof. Dr. Fábio Steiner – UEMS
- Prof. Dr. Gabriel Andres Tafur Gomez (Colômbia)
- Prof. Dr. Hebert Hernán Soto Gonzáles – UNAM (Peru)
- Prof. Dr. Hudson do Vale de Oliveira – IFRR
- Prof. Msc. Javier Revilla Armesto – UCG (México)
- Prof. Msc. João Camilo Sevilla – Mun. Rio de Janeiro
- Prof. Dr. José Luis Soto Gonzales – UNMSM (Peru)
- Prof. Dr. Julio Cezar Uzinski – UFMT
- Prof. Msc. Lucas R. Oliveira – Mun. de Chap. do Sul
- Prof. Dr. Leandro Argente-Martínez – ITSON (México)
- Profa. Msc. Lidiene Jaqueline de Souza Costa Marchesan – Consultório em Santa Maria
- Prof. Msc. Marcos Pisarski Júnior – UEG
- Prof. Dr. Mario Rodrigo Esparza Mantilla – UNAM (Peru)
- Profa. Msc. Mary Jose Almeida Pereira – SEDUC/PA
- Profa. Msc. Nila Luciana Vilhena Madureira – IFPA
- Profa. Dra. Patrícia Maurer
- Profa. Msc. Queila Pahim da Silva – IFB
- Prof. Dr. Rafael Chapman Auty – UO (Cuba)
- Prof. Dr. Rafael Felipe Ratke – UFMS
- Prof. Dr. Raphael Reis da Silva – UFPI

- Prof. Dr. Ricardo Alves de Araújo – UEMA
- Prof. Dr. Wéverson Lima Fonseca – UFPI
- Prof. Msc. Wesclen Vilar Nogueira – FURG
- Profa. Dra. Yilan Fung Boix – UO (Cuba)
- Prof. Dr. Willian Douglas Guilherme – UFT

Conselho Técnico Científico

- Esp. Joacir Mário Zuffo Júnior
- Esp. Maurício Amormino Júnior
- Esp. Tayronne de Almeida Rodrigues
- Esp. Camila Alves Pereira
- Lda. Rosalina Eufrausino Lustosa Zuffo

Ficha Catalográfica

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)	
A642	<p>Aplicações de inteligência artificial e ciência de dados [recurso eletrônico] / Organizadores Julio Cezar Uzinski, Caio Cesar Ensede de Abreu, Bruno Rodrigues de Oliveira. – Nova Xavantina, MT: Pantanal, 2020. 100p.</p> <p>Formato: PDF Requisitos de sistema: Adobe Acrobat Reader Modo de acesso: World Wide Web ISBN 978-65-88319-26-0 DOI https://doi.org/10.46420/9786588319260</p> <p>1. Inteligência artificial. 2. Automação. 3. Processamento de dados. I. Uzinski, Julio Cezar. II. Abreu, Caio Cesar Ensede de. III. Oliveira, Bruno Rodrigues de.</p> <p style="text-align: right;">CDD 001.535</p>
Elaborado por Maurício Amormino Júnior – CRB6/2422	

O conteúdo dos livros e capítulos, seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva do(s) autor (es). O download da obra é permitido e o compartilhamento desde que sejam citadas as referências dos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Pantanal Editora

Rua Abaete, 83, Sala B, Centro. CEP: 78690-000.
Nova Xavantina – Mato Grosso – Brasil.
Telefone (66) 99682-4165 (Whatsapp).
<https://www.editorapantanal.com.br>
contato@editorapantanal.com.br

APRESENTAÇÃO

Ciência de dados (*Data science*) e Inteligência artificial (*Artificial intelligence*) podem ser consideradas como algumas das áreas da ciência mais importantes de nosso tempo. Ademais, não estão apenas no mundo acadêmico ou no futuro como no filme *The terminator*, estão ao nosso redor nesse exato momento em situações simples, por exemplo, são empregadas para sugerir qual o próximo emoji que você enviará em uma conversa no seu aplicativo de bate papo favorito. Ciência de dados se refere a um campo de estudo muito amplo que se vale de métodos científicos para se obter informações e percepções de conjuntos de dados, enquanto que a inteligência artificial refere-se ao que chamamos de inteligência das máquinas. As aplicações dessas ferramentas estão relacionadas às tecnologias que nos envolvem em nosso cotidiano, e.g., as redes sociais ou comércio eletrônico, bem como aplicações que escapam ao senso comum, como em cibersegurança ou astrofísica.

Tais abordagens, às vezes, são utilizadas juntas, de forma intercalada ou separadas. Uma pesquisa poderá contar com ambas as ferramentas ou apenas com uma delas. Nesse livro, o leitor encontrará aplicações diversas dessas ferramentas em diferentes perspectivas. Em seis capítulos são apresentados resultados de pesquisas de dezesseis pesquisadores sobre temas diversos que dão uma demonstração do poder da Inteligência Artificial e da Ciência de Dados.

O primeiro capítulo traz uma comparação dos resultados da Transformada Wavelet (WT) e do Filtro Savitzky-Golay (SG) na estimação da tendência em séries temporais. Os resultados são obtidos a partir de uma série temporal artificial e uma série de dados reais, i.e, a cotação da moeda americana (Dólar US\$) frente à moeda brasileira (Real R\$).

Considerando que a principal abordagem em Aprendizado de Máquinas é a construção de algoritmos que, por meio das experiências e aprendizado, possam ser melhorados automaticamente, o segundo capítulo desse livro traz uma avaliação de um algoritmo usando Redes Neurais Artificiais LSTM. Essa abordagem é justificada pela necessidade de reconhecimento de padrões de comportamento de presença de elementos em um ambiente para realizar a predição de ações no local por meio dos dados coletados fornecendo uma base para o desenvolvimento de um sistema doméstico inteligente.

Já no terceiro capítulo, uma metodologia para descrição de Casos de Uso (Uc) no desenvolvimento de sistemas de informação é apresentada. A motivação para a criação desta metodologia é que muitos UC's são descritos com pouca riqueza de detalhes, o que prejudica a qualidade dos sistemas de informação.

O quarto capítulo traz uma aplicação bastante diversa de todas as outras: um estudo dos aspectos e causas da obesidade por meio do algoritmo de agrupamento Farthest First para agrupar

peças com informações demográficas, socioeconômicas, de biotipo e de conjunto de ações observáveis recorrentes.

No quinto capítulo é desenvolvida uma aplicação utilizando regressão logística como um mapeamento do espaço de atributos não-linearmente separável baseado em um tipo especial de funções denominadas de Golden Wavelets. Nesse texto, os autores buscam tratar a matemática envolvida no método de aprendizado de máquina denominado de Regressão Logística e propor a utilização de funções wavelets para mapeamento do espaço de atributos.

Para encerrar esse e-book com chave de ouro, o sexto capítulo, apresenta uma classificação de cenas acústicas utilizando a Transformada Wavelet e Aprendizagem de Máquina. Os autores dedicam a redação à apresentação dos principais métodos de Classificação de Cenas Acústicas (CCA) e os principais atributos utilizados para esta classificação, exemplos de aplicações, e uma investigação adicional centrada na utilização da Transformada Wavelet Contínua (TWC).

Caríssimo(a) leitor(a), convidamos você à leitura de Aplicações de Inteligência Artificial e Ciência de Dados, e esperamos que a mesma seja um deleite ilimitado e que contribua de alguma forma com suas aspirações.

Julio Cezar Uzinski


SUMÁRIO

Apresentação	4
Capítulo I	7
Obtenção de Tendências em Série Temporal empregando a Transformada Wavelet e o filtro Savitzky-Golay.....	7
Capítulo II	24
Aplicação de Redes Neurais Artificiais LSTM em domótica inteligente	24
Capítulo III	32
Uma Metodologia Simples para Descrição de Casos de Uso no Desenvolvimento de Sistemas de Informação.....	32
Capítulo IV	49
Aspectos e causas da obesidade: uma análise utilizando algoritmo de agrupamento <i>Farthest First</i>	49
Capítulo V	59
Mapeamento de espaço de atributos não-linearmente separável baseado na Golden wavelet: aplicação na regressão logística	59
Capítulo VI	85
Classificação de Cenas Acústicas Utilizando a Transformada Wavelet Contínua e Aprendizagem de Máquina	85
Índice Remissivo	99

Aspectos e causas da obesidade: uma análise utilizando algoritmo de agrupamento *Farthest First*

Recebido em: 05/08/2020


Aceito em: 17/08/2020

 10.46420/9786588319260cap4

Gilberto de Melo Junior^{1*} 

Sanderson Oliveira de Macedo¹ 

Renato Milhomem de Oliveira Filho² 

Jane Sousa Naves³ 

Géssica Helen de Melo⁴ 

INTRODUÇÃO

Muitos fatores comportamentais acarretam a obesidade em um indivíduo, isto é, há inúmeros fatores variáveis envolvidos. Portanto, a tarefa de classificar quais grupos de comportamentos são mais prevalentes no fator de desenvolvimento da obesidade é muito complexa (Ferreira e Magalhães, 2006). A alimentação descomedida não implica certa e diretamente em obesidade, pois existe um conjunto de fatores demográficos, socioeconômicos e biológicos que são determinantes para tal condição.

A obesidade representa um estado de alto e extenso aumento de gordura corporal total, sendo considerada como um problema mundial de saúde. Ela é caracterizada como uma condição complexa de ordem multifatorial, que leva ao surgimento ou acréscimo de doenças, estando a obesidade abdominal intimamente relacionada com o aumento significativo de doenças cardiovasculares e de ordem metabólica, como a hipertensão, dislipidemia e hiperglicemia (De Oliveira, 2017).

A obesidade é considerada uma doença do grupo de doenças crônicas não transmissíveis, com associações positivas para outros problemas de saúde como os cardiovasculares e cerebrovasculares, distúrbios metabólicos, diferentes tipos de câncer, doenças do aparelho digestivo, entre outras (Rocha-Brischiliari, 2014; Ducan, 2005) que são associadas principalmente com um conjunto de fatores demográficos, socioeconômicos e biológicos (Coelho, 2009).

Monteiro et al. (2003) explicam que as condições socioeconômicas, influenciam os hábitos de vida da população e, conseqüentemente o maior risco de obesidade. Fatores como: educação, renda e

¹ Docente do Departamento IV no Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

² Discente do curso de Engenharia de Controle e Automação no Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

³ Discente do curso de Biomedicina na Universidade Federal de Goiás.

⁴ Discente do curso de Medicina na Faculdade Morgana Potrich, Campus Mineiros.

* Autor correspondente: gilberto.melo@outlook.com

ocupação influenciam nos hábitos comportamentais que, por sua vez, afetam diretamente a ingestão maior de calorias e o gasto energético.

O aumento da prevalência de obesidade entre os jovens foi um dos principais temas de investigação da Saúde Pública na última década. Fundamentado nas observações de maior probabilidade do aparecimento de manifestações clínicas entre os jovens com excesso de peso corporal, a obesidade é considerada um grave fator de risco para morbidade na vida adulta e caracterizada, atualmente, como um dos grandes problemas de Saúde Pública (Cavazzotto et. al., 2014).

Assim, para determinar quais grupos estão mais propensos a serem obesos optou-se por utilizar a teoria de Aprendizagem de Máquina (*Machine Learning*) empregando a abordagem de aprendizagem não-supervisionada e a técnica de Agrupamentos. O objetivo desta técnica é encontrar uma estrutura de *clusters* (grupos) nos dados, onde os objetos pertencentes a cada cluster compartilham alguma característica ou propriedade relevante para o domínio do problema em estudo, ou seja, são de alguma maneira similares (Faceli et. al, 2011).

Dentre os trabalhos relacionados ao uso de *Machine Learning* aplicados a Nutrição, podemos citar: Freddo et al. (2019), onde os autores utilizaram o algoritmo de classificação de Árvore de Decisão para analisar 1339 prontuários de pacientes da Clínica Escola de Nutrição da UFFS. A acurácia apresentada no experimento foi superior de 89% na detecção de doenças relacionadas a alimentação. Já em Ramos et al. (2012), os autores utilizam algoritmos de Árvore de Decisão de *Machine Learning* para avaliação nutricional em crianças. Jacomini et al. (2018) utilizam algoritmos de Árvore de Decisão para obter padrões referentes a obesidade e seus fatores de risco cardiovascular e de comorbidade na cidade de Cascavel, Paraná, Brasil.

Nesse trabalho utilizou-se a técnica de aprendizagem de máquina não-supervisionada de clusterização (agrupamentos) *Farthest First*, para agrupar pessoas com informações demográficas, socioeconômicas, de biotipo e de conjunto de ações observáveis recorrentes, com o objetivo de criar uma base de referência para a análise e identificação de novos usuários com comportamentos tendenciais a obesidade. Essa abordagem resultou em uma nova ferramenta que poderá auxiliar profissionais da saúde e cientistas a identificarem pessoas que tem tendências à obesidade.

MATERIAL E MÉTODOS

A coleta de dados sobre os fatores que influenciam na determinação da obesidade, foi realizada utilizando um questionário online disponibilizado em um curto período de tempo (aproximadamente 1 mês) aos alunos do Instituto Federal de Educação, Ciência e Tecnologia de Goiás (IFG) do campus Goiânia e seus familiares, a fim de obter informações sobre suas rotinas diárias. Os participantes

concordaram com os termos da pesquisa e ficaram cientes que as respostas não seriam divulgadas, mas apenas utilizadas para fins científicos.

O questionário apresentou 29 questões, das quais todas foram elaboradas e discutidas com profissionais da área de saúde. Por meio dos dados coletados no questionário foi possível também calcular o Índice de Massa Corporal (IMC) de cada indivíduo, sendo o mesmo utilizado também como dado na clusterização.

O IMC é uma medida internacional usada para identificar se uma pessoa está no peso ideal ou não. Ele é determinado pela razão entre a massa do indivíduo e o quadrado da sua altura, onde a massa está em quilogramas e a altura é dada em metros (Anjos, 1992).

O questionário utilizado na coleta de dados apresentou as seguintes questões:

- Qual a sua idade?
- Qual a sua altura?
- Qual o seu peso em kg?
- Qual seu estado civil? (a-Solteiro(a) b-Casado(a) c-Divorciado(a) d-Viúvo(a) e- Namorando);
- Qual o seu meio de transporte? (a-Particular b-Público c-Táxi d- A Pé e-Bicicleta);
- Possui filhos? Quantos?
- Qual o seu nível de escolaridade? (a-Ensino Fundamental incompleto b- Ensino Fundamental Completo c-Ensino Médio Incompleto d-Ensino Médio Completo e- Ensino Superior Incompleto f-Ensino Superior Completo g-Especialização);
- Trabalha? Com o que? (a-Escritório/Professor b-Trabalho Pesado c-Domestica/Auxiliar de Serviços Gerais d-Vendedor/Andarilho e-Não Trabalha);
- Qual o seu nível salarial (ou outra remuneração)?
- Quantas refeições você faz por dia?
- Qual a quantidade (em média e em gramas) por refeição?
- Onde você faz suas principais refeições? (a- Casa b-Restaurante c- Leva a Própria Comida d-Casa e Restaurante);
- Consome bebida alcoólica? (a-Quase Nunca b-Socialmente c-Socialmente Exagerado d-Exagerado e-Não Consome);
- Têm o hábito de fumar? (a-Quase Nunca b-Socialmente c-Socialmente Exagerado d-Exagerado e-Não Fuma);

- Pratica alguma atividade física? Quantas vezes por semana? (a-1 b-2 c-3 d-4 e-5 f-6 g-7 h- Não pratico);
- Qual horário do dia? (a- Pela manhã b- Pela tarde c- Pela noite d- Não Pratico);
- Qual exercício físico? (a- Aeróbicos b- Anaeróbicos c- Ambos d- Não Pratico);
- Faz exames de rotina? (a- Sim b-Não);
- Em sua família, existem casos de: (a- Obesidade b- Hipertensão c- Diabetes d-Colesterol Elevado e-Triglicérides Elevado f- Mais de uma opção g- Nenhuma das opções);
- Em média, quantas horas dorme por dia?
- Você se considera: (a-Relaxado(a) b-Paciente c-Extrovertido(a) d-Calmo(a) e- Estressado(a) f- Bipolar g- Ansioso(a) h- Depressiva);
- Toma remédio controlado? (a- Sim b-Não);
- Já fez, ou faz, alguma dieta? (a- Sim b- Às Vezes c- Não);
- Já tomou ou toma *shakes* pra emagrecer ou engordar? (a- Sim b-Não);
- Já tomou ou toma remédio para emagrecer? (a- Manipulado b- Natural (chás, alimentação controlada) c- Farmácia d- Nunca Tomou);
- Quantos copos de água você consome por dia?
- Já fez alguma das seguintes cirurgias? (a-Bariátrica b-Tireoide c-Lipoaspiração d-Bariátrica e Lipoaspiração e-Bariátrica e Tireoide f-Tireoide e Lipoaspiração g-Nenhuma);
- Você possui alguma dessas doenças? (a- Hipertensão b- Colesterol Elevado c- Diabetes d- Triglicérides Elevado e- Mais de uma opção f-Nenhuma das opções);
- Possui vida sexual ativa? (a- Sim b-Não);
- Com qual frequência utiliza dispositivos eletrônicos? (celulares, computadores, etc.);
- Como você se vê de acordo com os corpos apresentado na Figura 1?

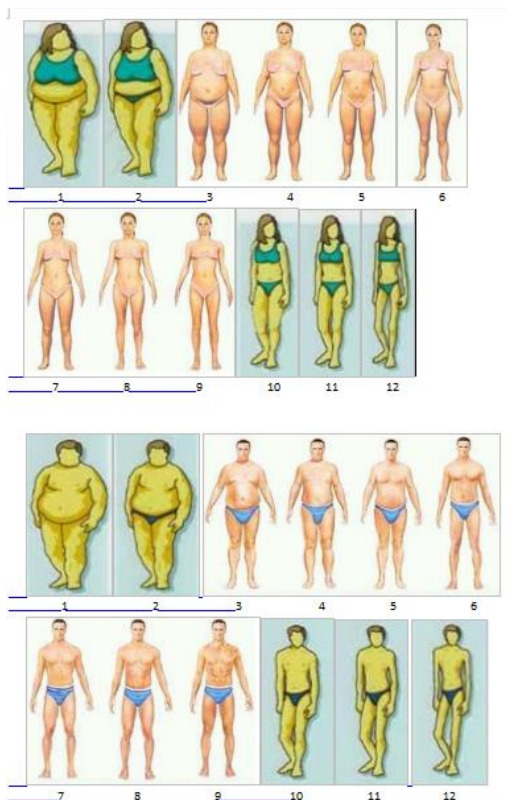


Figura 1. Estereótipos de corpos para o entrevistado escolher com qual se assemelha mais. Adaptado de Cataneo (2005).

ALGORITMO DE AGRUPAMENTO *FARTHEST FIRST*

O algoritmo de agrupamentos utilizado neste trabalho foi o *Farthest First*, que se configura como um algoritmo rápido e guloso. Ele foi proposto por Hochbaum e Shmoysat em 1985 (Vadegar e Yogish, 2014).

Neste algoritmo, k pontos são selecionados como os centros dos clusters. O primeiro centro é selecionado aleatoriamente. O segundo centro é avidamente selecionado como o ponto mais distante do primeiro. Cada centro restante é determinado selecionando o ponto mais distante do conjunto de objetos já escolhidos. Os pontos restantes são adicionados ao cluster cujo centro é o mais próximo (Kumar, 2013).

Ele possui complexidade de tempo $O(nk)$, onde n é o número de objetos no conjunto de dados e k é o número de clusters desejados (Dharmarajan e Velmurugan, 2015).

RESULTADOS E DISCUSSÃO

Foram respondidos 300 questionários, porém, uma averiguação destes mostrou que haviam respostas dados incompletos ou com dados errôneos (sem relação entre as perguntas). Assim, após o descarte desses dados restaram 250 respostas que foram utilizadas para comporem os dados de análises.

Com esses questionários calculou-se o índice de IMC de cada indivíduo, como informação que foi apresentada ao algoritmo para clusterização. Com um trabalho de adequação aos dados, os mesmos foram transformados em arquivo “.arff”. Este tipo de arquivo é utilizado para ser executado no software de código livre de mineração de dados Weka (Hall, 2009), utilizado neste trabalho.

Foi definido a criação de 5 clusters, com a perspectiva de alcançar os 5 grupos definidos pelo IMC (Magro, Normal, Médio, Gordo e Obeso). Embora o cálculo do IMC apresente algumas desvantagens e ele não tenha sido usado para clusterização, o mesmo possibilitou um melhor entendimento visual dos resultados aplicados. Utilizou-se o algoritmo no modo “Classe para avaliação do Cluster”. Neste modo o software inicialmente ignora o atributo de classe e gera o agrupamento. Em seguida, durante a fase de teste, ele atribui classes aos clusters, com base no valor majoritário do atributo de classe em cada cluster. A Tabela 1 apresenta os clusters que o algoritmo *Farthest First* analisou e criou utilizando os dados obtidos com a aplicação do questionário.

Tabela 1. Clusters obtidos após a clusterização utilizando o algoritmo *Farthest First*, juntamente com a porcentagem de cada cluster em relação ao total dos dados e também as respostas de cada pergunta que auxiliou na formação do cluster. Fonte: os autores.

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
22	47	22	18	17
1.53	1.87	1.63	1.89	1.57
49	125	76	65	64
Casado(a)	Casado(a)	Solteiro(a)	Solteiro(a)	Solteiro(a)
Particular	A pé	Particular	Público	A pé
0	2	0	0	0
Superior	Especialização	Superior	Ensino Médio	Ensino Técnico
Não trabalha	Professor /	Não trabalha	Não trabalha	Professor /
400	20000	0	0	450
3	3	6	5	3
350	500	150	500	800
Leva a própria	Casa	Casa	Casa e	Casa
Não consome	Socialmente	Socialmente	Socialmente	Não consome
Não fuma	Não fuma	Quase nunca	Não fuma	Não fuma
3	Não pratica	5	2	Não pratica
A noite	Não pratica	Pela manhã	Pela tarde	Não pratica
Aeróbicos	Não pratica	Ambos	Ambos	Não pratica
Sim	Não	Sim	Não	Não
Mais de uma	Obesidade	Mais de uma	Diabetes	Hipertensão
7	8	7	9	6
Ansioso(a)	Bipolar	Extrovertido(a)	Calmo(a)	Calmo(a)
Não	Sim	Não	Não	Não
Não	Sim	Sim	Não	Não
Não	Não	Sim	Não	Sim

Nunca tomou	Nunca tomou	Todos	Nunca tomou	Natural
3	4	8	10	5
Nenhuma	Nenhuma	Bariátrica	Nenhuma	Nenhuma
Colesterol	Triglicérides	Nenhuma das	Nenhuma das	Colesterol elevado
Sim	Sim	Não	Não	Não
3	8	16	15	4
6	4	5	11	4
20.13	35.74	28.6	18.19	25.96
Normal	Obeso	Gordo	Magro	Médio

O Cluster 0 definiu o grupo denominado normal, para 44% dos participantes. As características desse grupo foram mulheres, com idade de 20 anos, solteiras, com nível superior incompleto, que mantém hábitos diários considerados saudáveis como: alimentação fracionada e equilibrada, acompanhamento médico regular, além da prática diária de atividades físicas. Apesar dessas participantes terem respondido como afirmativa a presença de doenças crônicas em suas famílias, como hipertensão, diabetes e obesidade, elas não confirmaram a presença de tais diagnósticos para si mesmas. Porém, é importante observar que o IMC dessas participantes foi de 24,6, o que indica que a obesidade está relacionada, de uma forma geral, a estilos de vida e hábitos alimentares.

O Cluster 1 identificou 12% dos participantes propensos ao desenvolvimento da obesidade, denominados pela Organização Mundial de Saúde como obesidade grau 1 ($IMC \geq 30$). Neste grupo, prevaleceu o sexo masculino, com idade de 21 anos. Os participantes negaram a prática regular de atividades físicas, ingestão de altas quantidades de alimentos com intervalos superiores a 3 horas, uso de bebidas alcoólicas e ausência de acompanhamento médico de rotina.

De acordo com Cataneo (2005), a condição socioeconômica, influencia os hábitos de vida da população e, conseqüentemente o risco de obesidade. Fatores, como educação, renda e ocupação influenciam nos hábitos comportamentais que afetam diretamente a ingestão calórica e gasto energético. Assim, quando alimentos saudáveis estão menos disponíveis à população de baixa renda, como peixes, grãos, carnes magras, frutas e verduras, observa-se um aumento da relação entre obesidade e nível socioeconômico. É o que também pode ser analisado pelos resultados do cluster 1. Neste grupo, a renda de 1 salário mínimo, o nível de escolarização e a rotina diária, colaboram para um padrão comportamental propenso ao desenvolvimento da obesidade.

O Cluster 2 mostrou que 7% dos participantes estão abaixo do grupo de risco para sobrepeso, com $IMC < 29$. Os participantes desse grupo são do sexo masculino, com idade de 25 anos e nível de escolaridade em nível de especialização. Neste grupo, observou-se a presença de doenças como o diabetes e triglicérides elevados, uso de bebidas alcoólicas e prática esporádica de atividades físicas.

O Cluster 3 definiu que 34% dos participantes são mulheres, com idade de 43 anos, solteiras, com renda salarial aproximada de 7 salários mínimos e nível superior completo. Em suas rotinas, destaca-se a prática de atividades físicas 2 vezes por semana e a ausência do uso de álcool e tabaco.

Fatores psicológicos como passividade, submissão, dependência, preocupação excessiva com comida, ingestão compulsiva de alimentos, baixa aceitação do esquema corporal, dificuldades de adaptação social, medo de rejeição, sentimento de desamparo, intolerância e culpa podem estar associados a comportamentos de adultos obesos que fazem hiperfagia. O ato de comer, para pessoas ansiosas e instáveis emocionalmente pode ser considerado como um tranquilizador, quando há dificuldades de lidar com a frustração e limites. Assim, esses fatores, integram um quadro multifatorial que pode predispor o sujeito à obesidade, juntamente com outros fatores como os genéticos e ambientais (Cataneo, 2005). É o que verificamos no Cluster 4, em que houve uma prevalência maior de participantes com histórico familiar de obesidade, triglicérides elevadas, ausência de atividades físicas e problemas psicológicos, como o transtorno do humor bipolar, que tem como pano de fundo, um funcionamento de personalidade do tipo ansioso. Adicionalmente, 3% dos participantes do sexo masculino, com idade de 47 anos, casados, com nível superior em nível de pós-graduação e renda salarial aproximada de 3 salários mínimos, afirmaram não praticarem atividades físicas, fazerem uso esporádico de álcool e apresentarem na família problemas de saúde como obesidade.

CONCLUSÃO

Conclui-se que neste estudo onde o objetivo foi a aplicação e análise de aspectos ligados a obesidade utilizando uma técnica de aprendizagem de máquina para agrupar informações de pessoas com características demográficas, socioeconômicas, de biotipo e de comportamento recorrente que foram obtidas por meio de um questionário online. Utilizou-se o algoritmo de aprendizagem não-supervisionado *Farthest First* que analisou os dados e criou cinco agrupamentos de indivíduos, com o objetivo de ser uma base de referência para a análise e identificação de novos usuários com comportamentos que tendem a obesidade. Verificou-se que a utilização desse instrumento pode avaliar satisfatoriamente o padrão comportamental e identificar fatores de risco associados ao desenvolvimento da obesidade. A organização dos dados em clusters permitiu uma visualização das características gerais da população pesquisada, oferecendo informações úteis para a prevenção dos riscos da obesidade e suas comorbidades.

Em trabalhos futuros pretende-se analisar tais grupos utilizando outros algoritmos de aprendizagem não-supervisionada, ou ainda organizando os dados para aplicação em algoritmos supervisionados onde poderemos obter modelos preditivos de classificação.

AGRADECIMENTOS

Os autores agradecem o suporte do Instituto Federal de Educação, Ciência e Tecnologia de Goiás (IFG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERÊNCIAS BIBLIOGRÁFICAS

- Anjos, LA (1992). Índice de massa corporal (massa corporal. estatura-2) como indicador do estado nutricional de adultos: revisão da literatura. *Revista de Saúde pública*, 26: 431-436.
- Cataneo C, Carvalho AMP, Galindo EMC (2005). Obesidade e aspectos psicológicos: maturidade emocional, auto-conceito, locus de controle e ansiedade. *Psicologia: Reflexão e crítica*, 18(1): 39-46.
- Cavazzotto, TG, Brasil, MR, Oliveira, VM, Silva, SRD, Ronque, ERV, Queiroga, MR, & Serassuelo Junior, H (2014). Nutritional status of children and adolescents based on body mass index: agreement between World Health Organization and International Obesity Task Force. *Revista Paulista de Pediatria*, 32(1): 44-49.
- Coelho, MSPH, MOURA, ECD, & ASSIS, MAAD (2009). Aumento do índice de massa corporal após os 20 anos de idade e associação com indicadores de risco ou de proteção para doenças crônicas não transmissíveis. Monteiro, CA, Conde, WL, Castro, IRRD. A tendência cambiante da relação entre escolaridade e risco de obesidade no Brasil (1975-1997). 2013, *Cadernos de Saúde Pública*, 19: S67-S75.
- Dharmarajan, A., & Velmurugan, T. (2015). Lung cancer data analysis by k-means and farthest first clustering algorithms. *Indian Journal of science and Technology*, 8(15): 1-8.
- de Oliveira Couto, J, Santos, WS, Santos, JR, Simões, AEC, dos Santos, ÉGM, & dos Santos Silva, RJ (2017). Risco cardiovascular, índices antropométricos e percepção de qualidade de vida em idosos. *Scientia Plena*, 13(3).
- Duncan, BB, Duncan, MS, & Schmidt, MI (2005). Inflamação subclínica, obesidade, diabetes e doenças relacionadas. *Revista HCPA*, 25(3): 5-16.
- Faceli K, Lorena AC, Gama J, Carvalho ACPLF de (2011). *Uma Abordagem de Aprendizagem de Máquina*. Rio de Janeiro: LTC.
- Ferreira VA, Magalhães R (2006). Obesidade no Brasil: tendências atuais. *Revista portuguesa de saúde pública*, 24(2): 71-81.
- Freddo AR, Nishiyama MF, Zanuzo K, Koehnlein E (2019). Árvores de Decisão como Método de Mineração de Dados: Análise de Prontuários de uma Clínica Escola de Nutrição. *Revista da Associação Brasileira de Nutrição*, 10(2): 31-37.

- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1): 10-18.
- Jacomini D, de Bona EM, Johann J (2018). Mineração de dados aplicada aos riscos da obesidade em Cascavel/PR. *Extensio: Revista Eletrônica de Extensão*, 15(30): 26-38.
- Kumar M (2013, November). An optimized farthest first clustering algorithm. In 2013 Nirma University International Conference on Engineering (NUICONE) (pp. 1-5). IEEE.
- Monteiro CA, Conde WL, Castro IRRD (2003). A tendência cambiante da relação entre escolaridade e risco de obesidade no Brasil (1975-1997). *Cadernos de Saúde Pública*, 19: S67-S75.
- Ramos RM, Palmero RMR, Ávalos RG, Lorenzo MMG (2012). Aplicación de métodos de selección de atributos para determinar factores relevantes en la evaluación nutricional de los niños. *Gaceta Médica Espirituana*, 9(1): 7.
- Rocha-Brischiliari SC, Agnolo CMD, Gravena AAF, Lopes TCR, Carvalho MDB, Pelloso SM Doenças crônicas não transmissíveis e associação com fatores de risco; 2014. *Rev Bras Cardiol*, 27(1): 35-42.
- Vadeyar DA, Yogish HK (2014). Farthest first clustering in links reorganization. *International Journal of Web & Semantic Technology*, 5(3): 17.

ÍNDICE REMISSIVO

A

agrupamento, 49, 53, 54
aprendizado de máquina, 8, 25, 28, 29, 59, 60, 61, 80
Aprendizagem de Máquina, 50, 57, 83

C

caso de uso, 41, 42, 44, 46
CCA, 84, 85, 86, 87, 88, 95
cenário, 39, 43, 44, 45, 85
Cenas Acústicas, 83, 84, 95

Ch

checklist, 34, 39, 41

C

Ciência de Dados, 33
classificação, 31, 39, 50, 56, 59, 60, 61, 62, 63, 68, 69, 72, 73, 74, 84, 85, 88, 90, 94
clusterização, 50, 51, 54
coeficiente, 10, 11, 89, 94
comportamento, 8, 9, 22, 26, 30, 32, 37, 56, 70

D

Deep Learning, 31, 32, 59, 81, 87, 90, 94
desenvolvimento de SI, 35, 36, 39
domótica, 25, 26

E

Engenharia de Requisitos, 33, 47, 48
escalograma, 89, 90
espaço de atributos, 59, 60, 70, 73, 74, 76, 80, 81
espectro, 86, 89
espectrograma, 85, 86

F

Farthest First, 49, 50, 53, 54, 56

filtro, 8, 9, 10, 11, 15, 16, 17, 18, 19, 20, 21, 22, 89

G

GWS, 89, 90, 91, 92, 93, 94

I

Inteligência Artificial, 25

M

machine learning, 30, 31, 84, 87
mapeamento, 48, 60, 70, 73, 74, 77, 78, 80, 81, 88
metodologia, 33, 34, 35, 36, 37, 39, 41, 42, 47
MFCCs, 84, 86, 87, 90, 91, 93, 94

O

obesidade, 49, 50, 55, 56, 57, 58

P

polinômio, 10, 11, 16, 20, 21
prototipação, 33, 35, 39, 41, 47

Q

qualidade, 33, 34, 35, 39, 41, 47, 57

R

Redes Neurais
 Artificiais, 25, 26
 LSTM, 26
regressão logística, 59, 62, 63, 70
requisitos, 28, 33, 34, 35, 39, 41, 48

S

Savitzky-Golay, 8, 9, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23
sensores, 26, 27, 29
Série Temporal, 8, 9
Sistema de Informação, 47, 48

Software, 47, 48, 95

superfície de decisão, 62, 77

T

tendência, 8, 9, 10, 14, 15, 16, 17, 18, 19, 21,
22, 57, 58

Transformada Wavelet, 8, 9, 10, 13, 70, 83, 84,
87, 88, 90, 95



Pantanal Editora

Rua Abaete, 83, Sala B, Centro. CEP: 78690-000
Nova Xavantina – Mato Grosso – Brasil
Telefone (66) 99682-4165 (Whatsapp)
<https://www.editorapantanal.com.br>
contato@editorapantanal.com.br

ISBN 978-658831926-0

