09/09/2022 870220082323 16:29

29409161948459603

Pedido nacional de Invenção, Modelo de Utilidade, Certificado de Adição de Invenção e entrada na fase nacional do PCT

Número do Processo: BR 10 2022 018103 9

Dados do Depositante (71)

Depositante 1 de 2

Nome ou Razão Social: UNIVERSIDADE FEDERAL DE MINAS GERAIS

Tipo de Pessoa: Pessoa Jurídica

CPF/CNPJ: 17217985000104

Nacionalidade: Brasileira

Qualificação Jurídica: Instituição de Ensino e Pesquisa

Endereço: Av. Antônio Carlos, 6627 - Unidade Administrativa II - 2º andar- sala 2011

Cidade: Belo Horizonte

Estado: MG

CEP: 31270-901

País: Brasil

Telefone: (31) 3409-6430

Fax:

Email: patentes@ctit.ufmg.br

PETICIONAMENTO ELETRÔNICO

Esta solicitação foi enviada pelo sistema Peticionamento Eletrônico em 09/09/2022 às 16:29, Petição 870220082323

Depositante 2 de 2

Nome ou Razão Social:	KUNUMI SERVIÇOS EM TECNOLOGIA DA INFORMAÇÃO S.A.
Tipo de Pessoa:	Pessoa Jurídica
CPF/CNPJ:	24477718000131
Nacionalidade:	Brasileira
Qualificação Jurídica:	Pessoa Jurídica
Endereço:	Rua Rio Grande do Norte, 1.435, sala 708 (parte) 7o pavimento, bairro Savassi
Cidade:	Belo Horizonte
Estado:	MG
CEP:	30130-131
País:	BRASIL
Telefone:	
Fax:	
Email:	

Dados do Pedido

Natureza Patente: 10 - Patente de Invenção (PI)

Título da Invenção ou Modelo de Utilidade (54): PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES E USO

Resumo: Esta tecnologia refere-se a um processo de obtenção de modelos baseados em Aprendizado de Máquina que inclui meios para promover a diversificação das explicações do modelo. A tecnologia utiliza técnicas como a decomposição do espaço de dados em estruturas locais, do tipo backbone, formando um conjunto do tipo backbone features; representações de explicações de modelos (explicabilidade) utilizadas como critério de clusterização e seleção de protótipos dentro do espaço de modelos para promover a diversificação de explicações. Utiliza também combinação (ensemble) de modelos, dentre outras técnicas. As vantagens propiciadas pela tecnologia são principalmente: 1) agilidade na modelagem devido à criação de características relevantes ao problema, economizando tempo e recursos computacionais, favorecendo os ajustes de hiperparâmetros; 2) versatilidade, já que a metodologia pode ser utilizada em problemas de qualquer segmento e formato de banco de dados; 3) propicia o entendimento dos modelos devido à utilização de técnicas de explicabilidade; 4) viabiliza a diversificação das explicações do modelo. A tecnologia aplica-se no contexto de solução de problemas em que se recorre a modelos baseados em Aprendizado de Máquina, para aperfeiçoar a elaboração e o desempenho de tais modelos.

Figura a publicar: 1

Dados do Inventor (72)

Inventor 1 de 3

Nome: ADRIANO ALONSO VELOSO,

CPF: 03733647688

Nacionalidade: Brasileira

Qualificação Física: Professor do ensino superior

Endereço: Av. Antônio Carlos, 6627 - Unidade Administrativa II - 2º andar- sala 2011

Cidade: Belo Horizonte

Estado: MG

CEP: 31270-901

País: BRASIL

Telefone: (31) 340 93932

Fax:

Email: patentes@ctit.ufmg.br

Inventor 2 de 3

Nome: ANDERSON BESSA DA COSTA,

CPF: 02275151109

Nacionalidade: Brasileira

Qualificação Física: Pesquisador

Endereço: Av. Antônio Carlos, 6627 - Unidade Administrativa II - 2º andar- sala 2011

Cidade: Belo Horizonte

Estado: MG

CEP: 31270-901

País: BRASIL

Telefone: (31) 340 93932

Fax:

Email: patentes@ctit.ufmg.br

Inventor 3 de 3

Nome: NIVIO ZIVIANI

CPF: 07230257620

Nacionalidade: Brasileira

Qualificação Física: Professor do ensino superior

Endereço: Av. Pres. Antônio Carlos, 6627 - Pampulha- 2º andar- sala

Cidade: Belo Horizonte

Estado: MG

CEP: 31270-901

País: BRASIL

Telefone: (31) 340 93932

Fax:

Email: patentes@ctit.ufmg.br

Documentos anexados

Tipo Anexo	Nome
Comprovante de pagamento de GRU 200	1 - Depósito PI - 29409161948459603.pdf
Portaria	2 - Portaria 2195-2020 - Prof. Gilberto UFMG.pdf
Procuração	3 - Procuração para UFMG - NI 49.2021.docx - Clicksign.pdf
Comprovação de Poderes	4 - Comprovação de Poderes.pdf
Relatório Descritivo	5 - RELATÓRIO DESCRITIVO.pdf
Reivindicação	6 - REIVINDICAÇÕES.pdf
Desenho	7 - DESENHOS.pdf
Resumo	8 - RESUMO.pdf
Declaração de período de graça	9 - Período de Graça NI 49 - 2021.pdf

Acesso ao Patrimônio Genético

☑ Declaração Negativa de Acesso - Declaro que o objeto do presente pedido de patente de invenção não foi obtido em decorrência de acesso à amostra de componente do Patrimônio Genético Brasileiro, o acesso foi realizado antes de 30 de junho de 2000, ou não se aplica.

Declaração de Divulgação Anterior Não Prejudicial

☑ Artigo 12 da LPI - Período de Graça.

Declaração de veracidade

☑ Declaro, sob as penas da lei, que todas as informações acima prestadas são completas e verdadeiras.

__ SIAFI2022-DOCUMENTO-CONSULTA-CONGRU (CONSULTA GUIA DE RECOLHIMENTO DA UNIAO
07/04/22 16:48   USUARIO : LUDMILA
DATA EMISSAO : 07Abr22   TIPO : 1 - PAGAMENTO   NUMERO : 2022GR800160
UG/GESTAO EMITENTE   : 153254 / 15229 - ADMINISTRACAO GERAL/UFMG
UG/GESTAO FAVORECIDA : 183038 / 18801 - INSTITUTO NACIONAL DA PROPRIEDADE INDU
RECOLHEDOR   : 153254   GESTAO : 15229
CODIGO RECOLHIMENTO : 72200 - 6 COMPETENCIA: ABR22   VENCIMENTO: 07Abr22
DOC. ORIGEM: 153254 / 15229 / 2022NP000569   PROCESSO :
RECURSO : 1
(=)VALOR DOCUMENTO   :   70,00
(-)DESCONTO/ABATIMENTO:
(-)OUTRAS DEDUCOES   :
(+)MORA/MULTA   :
(+)JUROS/ENCARGOS   :
(+)OUTROS ACRESCIMOS :
(=)VALOR TOTAL   :   70,00
NOSSO NUMERO/NUMERO REFERENCIA : 00029409161948459603
CODIGO DE BARRAS : 89610000000 0 70000001010 3 95523127220 9 00360640000 4
OBSERVACAO
Serviço: 200-Pedido nacional de Invenção, Modelo de Utilidade, Certificado de Adição de Invenção e entrada na fase nacional do PCT
LANCADO POR : 09663457627 - LUDMILA   UG : 153254   07Abr2022   16:40
PF1=AJUDA PF3=SAI PF2=DADOS ORC/FIN PF4=ESPELHO PF12=RETORNA

INSTRUÇÕES:

A data de vencimento não prevalece sobre o prazo legal. O pagamento deve ser efetuado antes do protocolo. Órgãos públicos que utilizam o sistema SIAFI devem utilizar o número da GRU no campo Número de Referência na emissão do pagamento. Serviço: 200-Pedido nacional de Invenção, Modelo de Utilidade, Certificado de Adição de Invenção e entrada na fase nacional do PCT

Clique aqui e pague este boleto através do Auto Atendimento Pessoa Física.

Clique aqui e pague este boleto através do Auto Atendimento Pessoa Jurídica.

Recibo do Pagador

BANCO DO BRASIL | **001-9** | **00190.00009 02940.916196 48459.603170 3 89740000007000**

Nome do Pagador/CPF/CNPJ/Endereço
UNIVERSIDADE FEDERAL DE MINAS GERAIS CPF/CNPJ: 17217985000104
AV ANTONIO CARLOS 6627 UNIDADE ADMINISTRATIVA II 2 ANDAR SALA 2011, BELO HORIZONTE -MG CEP:31270901
Sacador/Avalista

Nosso-Número	Nr. Documento	Data de Vencimento	Valor do Documento	(=) Valor Pago
29409161948459603	29409161948459603	03/05/2022	70,00	

Nome do Beneficiário/CPF/CNPJ/Endereço
INSTITUTO NACIONAL DA PROPRIEDADE INDUST CPF/CNPJ: 42.521.088/0001-37
RUA MAYRINK VEIGA 9 24 ANDAR ED WHITE MARTINS , RIO DE JANEIRO - RJ CEP: 20090910

Agência/Código do Beneficiário
2234-9 / 333028-1

Autenticação Mecânica

BANCO DO BRASIL | **001-9** | **00190.00009 02940.916196 48459.603170 3 89740000007000**

Local de Pagamento
PAGÁVEL EM QUALQUER BANCO ATÉ O VENCIMENTO

Data de Vencimento
03/05/2022

Nome do Beneficiário/CPF/CNPJ
INSTITUTO NACIONAL DA PROPRIEDADE INDUST CPF/CNPJ: 42.521.088/0001-37

Agência/Código do Beneficiário
2234-9 / 333028-1

Data do Documento	Nr. Documento	Espécie DOC	Aceite	Data do Processamento
04/04/2022	29409161948459603	DS	N	04/04/2022

Nosso-Número
29409161948459603

Uso do Banco	Carteira	Espécie	Quantidade	xValor
29409161948459603	17	R$		

(=) Valor do Documento
70,00

Informações de Responsabilidade do Beneficiário

A data de vencimento não prevalece sobre o prazo legal.
O pagamento deve ser efetuado antes do protocolo.
Órgãos públicos que utilizam o sistema SIAFI devem utilizar o número da GRU n
o campo Número de Referência na emissão do pagamento.
Serviço: 200-Pedido nacional de Invenção, Modelo de Utilidade, Certificado de
Adição de Invenção e entrada na fase nacional do PCT

(-) Desconto/Abatimento

(+) Juros/Multa

(=) Valor Cobrado

Nome do Pagador/CPF/CNPJ/Endereço
UNIVERSIDADE FEDERAL DE MINAS GERAIS CPF/CNPJ: 17217985000104
AV ANTONIO CARLOS 6627 UNIDADE ADMINISTRATIVA II 2 ANDAR SALA 2011,
BELO HORIZONTE-MG CEP:31270901

Código de Baixa

Sacador/Avalista

Autenticação Mecânica - **Ficha de Compensação**

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PORTARIA Nº 2195, DE 06 DE ABRIL DE 2020

A REITORA DA UNIVERSIDADE FEDERAL DE MINAS GERAIS, no uso de suas atribuições legais e estatutárias, considerando o disposto nos artigos 11 e 12 do Decreto-Lei nº 200, de 25 de fevereiro de 1967,

RESOLVE:

Art. 1º Delegar competência ao Diretor da Coordenadoria de Transferência e Inovação Tecnológica (CTIT), Professor Gilberto Medeiros Ribeiro, Inscrição UFMG nº 247405 e SIAPE nº 1964486, e a seu substituto eventual para, no âmbito desse Órgão,

a) assinar, por meio eletrônico ou físico, documentos ou instrumentos jurídicos, concernentes ao exercício das atividades de competência da CTIT, no âmbito da Lei 10.973/04 – Lei de Inovação Tecnológica, da Política de Inovação da UFMG e suas resoluções específicas, tais como Contrato de Transferência de *Know-How*, Contrato de Licenciamento de Tecnologia, Contrato de Partilhamento de Titularidade de Tecnologia, Acordos de Confidencialidade e Termos de Sigilo, Termos de Autorização de Teste e documentos afins;

b) assinar, por meio eletrônico ou físico, documentação necessária para depósito, processamento, adição, retificação, substituição, modificação, ampliação e resposta de relatórios referentes a objeto de proteção de propriedade intelectual junto aos órgãos competentes, em âmbito nacional e internacional;

c) autorizar a realização de despesas dentro dos limites orçamentários da CTIT;

d) autorizar a concessão de suprimento de fundos a servidores da Unidade, bem como determinar a baixa de responsabilidade;

e) requisitar passagens e transportes em geral, por quaisquer vias, nos limites da dotação orçamentária da CTIT;

f) autorizar viagens de servidores, a serviço da Unidade, arbitrando-lhes as respectivas diárias, obedecidas as disposições legais pertinentes;

g) assinar contratos, decorrentes de licitação, de dispensa de licitação ou inexigibilidade, no âmbito da CTIT;

h) prover arrecadação de receitas em geral, no âmbito da CTIT; e

i) apurar dívidas de terceiros para com a Universidade, oriundas de contratos de cotitularidade, licenciamento, transferência, dentre outros, adotando as medidas necessárias à regularização delas, no âmbito da CTIT.

Art. 2º Com base no disposto no Decreto nº 10.193, de 27 de dezembro de 2019, e no inciso II do art. 1º e art. 3º da Portaria nº 243, de 12 de fevereiro de 2020, do Ministério da Educação (MEC), subdelegar

competência ao supracitado Diretor e a seu substituto eventual para, no âmbito da CTIT,

I - celebrar novos contratos administrativos decorrentes de licitação, de dispensa de licitação e de inexigibilidade, ou prorrogar contratos em vigor relativos às atividades de custeio cujos valores sejam inferiores a R$500.000,00 (quinhentos mil reais); e

II - autorizar a realização de despesas relativas às atividades de custeio cujos valores sejam inferiores a R$500.000,00 (quinhentos mil reais).

Art. 3º Tornar sem efeito a Portaria nº 010, de 24 de janeiro de 2019.

Art. 4º A presente Portaria entra em vigor nesta data.

Belo Horizonte, 6 de abril de 2020.

Profa. Sandra Regina Goulart Almeida

Reitora

Documento assinado eletronicamente por **Sandra Regina Goulart Almeida**, **Reitora**, em 09/04/2020, às 17:17, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0096203** e o código CRC **04D898C8**.

Referência: Processo nº 23072.209054/2020-42 SEI nº 0096203

PROCURAÇÃO

Por este instrumento particular de Procuração, a KUNUMI SERVIÇOS EM TECNOLOGIA DA INFORMAÇÃO S.A. , sediada na Rua Rio Grande do Norte, 1.435, sala 708 (parte) 7o pavimento, bairro Savassi, na Cidade de Belo Horizonte, Estado de Minas Gerais, inscrita no CNPJ sob o n° 24.477.718/0001-31, neste ato representado por por Alberto Henrique Duarte Colares e Maurício Campos Zuardi, confere poderes especiais à **UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG**, com sede na Avenida Antônio Carlos, nº 6.627, Belo Horizonte, Minas Gerais, inscrita no CNPJ sob o nº 17.217.985/0001-04, representada neste ato pelo Professor Gilberto Medeiros Ribeiro, Diretor da Coordenadoria de Transferência e Inovação Tecnológica – CTIT, para representá-la perante o Instituto Nacional da Propriedade Industrial – INPI, para o fim de requerer e processar direitos de propriedade intelectual em cotitularidade entre a Kunumi e a UFMG face ao pedido de patente intitulado "PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES E USO", a ser depositado junto ao INPI, para mantê-lo em vigor com amplos poderes para assinar petições e documentos, pagar taxas, anotar transferências, fazer prova de uso da invenção patenteada, apresentar oposições, recursos, réplicas, anotar, elaborar notificações extrajudiciais, e praticar para os fins mencionados todos os atos necessários perante as autoridades administrativas competentes no Brasil e no exterior, em benefício da Outorgante, ratificando os atos já praticados.

Belo Horizonte/MG, 31 de agosto de 2022.

_____

Alberto Henrique Duarte Colares

_____

Maurício Campos Zuardi

2 Procuração para UFMG - NI 49.2021.docx.pdf

Documento número #f64d1c33-2bfe-4509-9c32-6c10738c2a12

Hash do documento original (SHA256): eb57e2356b0d5fa1a316a609e0452c45e6e69f8f25acf3f77fbd350ac285eb4d

Hash do PAdES (SHA256): 40e06afe2048157d52231648e4a2b83be60ad78c4db86f73412004a40483101f

Assinaturas

MAURICIO CAMPO ZUARDI

CPF: 260.677.618-66

Assinou como representante legal em 01 set 2022 às 10:52:57

Emitido por AC VALID RFB v5- com Certificado Digital ICP-Brasil válido até 11 out 2022

ALBERTO HENRIQUE DUARTE COLARES

CPF: 818.097.436-72

Assinou como representante legal em 01 set 2022 às 10:54:25

Emitido por AC LINK RFB v2- com Certificado Digital ICP-Brasil válido até 07 jun 2023

Log

31 ago 2022, 11:58:25	Operador com email marina.sasaki@kunumi.com na Conta f2eeac1c-c295-4a27-a8af-ad93f7bea379 criou este documento número f64d1c33-2bfe-4509-9c32-6c10738c2a12. Data limite para assinatura do documento: 30 de setembro de 2022 (11:55). Finalização automática após a última assinatura: habilitada. Idioma: Português brasileiro.
31 ago 2022, 11:58:27	Operador com email marina.sasaki@kunumi.com na Conta f2eeac1c-c295-4a27-a8af-ad93f7bea379 adicionou à Lista de Assinatura: mauricio@kunumi.com para assinar como representante legal, via E-mail, com os pontos de autenticação: Certificado Digital; Nome Completo; CPF; endereço de IP.
31 ago 2022, 11:58:27	Operador com email marina.sasaki@kunumi.com na Conta f2eeac1c-c295-4a27-a8af-ad93f7bea379 adicionou à Lista de Assinatura: alberto@kunumi.com para assinar como representante legal, via E-mail, com os pontos de autenticação: Certificado Digital; Nome Completo; CPF; endereço de IP.
01 set 2022, 10:52:58	MAURICIO CAMPO ZUARDI assinou como representante legal. Pontos de autenticação: certificado digital, tipo A3 e-cpf. CPF informado: 260.677.618-66. IP: 186.249.226.111. Componente de assinatura versão 1.353.0 disponibilizado em https://app.clicksign.com.
01 set 2022, 10:54:25	ALBERTO HENRIQUE DUARTE COLARES assinou como representante legal. Pontos de autenticação: certificado digital, tipo A1 e-cpf. CPF informado: 818.097.436-72. IP: 186.249.226.111. Componente de assinatura versão 1.353.0 disponibilizado em https://app.clicksign.com.

01 set 2022, 10:54:26	Processo de assinatura finalizado automaticamente. Motivo: finalização automática após a última assinatura habilitada. Processo de assinatura concluído para o documento número f64d1c33-2bfe-4509-9c32-6c10738c2a12.

Documento assinado com validade jurídica.
Para conferir a validade, acesse https://validador.clicksign.com e utilize a senha gerada pelos signatários ou envie este arquivo em PDF.
As assinaturas digitais e eletrônicas têm validade jurídica prevista na Medida Provisória nº. 2200-2 / 2001

Este Log é exclusivo e deve ser considerado parte do documento nº f64d1c33-2bfe-4509-9c32-6c10738c2a12, com os efeitos prescritos nos Termos de Uso da Clicksign, disponível em www.clicksign.com.

Ministério da Economia
Secretaria de Governo Digital
Departamento Nacional de Registro Empresarial e Integração
Secretaria de Estado de Fazenda de Minas Gerais

Nº DO PROTOCOLO (Uso da Junta Comercial)

NIRE (da sede ou filial, quando a sede for em outra UF)	Código da Natureza Jurídica	Nº de Matrícula do Agente Auxiliar do Comércio
31300114333	2054	

1 - REQUERIMENTO

ILMO(A). SR.(A) PRESIDENTE DA Junta Comercial do Estado de Minas Gerais

Nome: KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A
(da Empresa ou do Agente Auxiliar do Comércio)

Nº FCN/REMP
MGN2066110248

requer a V.Sª o deferimento do seguinte ato:

Nº DE VIAS	CÓDIGO DO ATO	CÓDIGO DO EVENTO	QTDE	DESCRIÇÃO DO ATO / EVENTO
1	007			ATA DE ASSEMBLEIA GERAL EXTRAORDINARIA
		219	1	ELEICAO/DESTITUICAO DE DIRETORES

BELO HORIZONTE
Local

27 Maio 2020
Data

Representante Legal da Empresa / Agente Auxiliar do Comércio:
Nome: _____
Assinatura: _____
Telefone de Contato: _____

2 - USO DA JUNTA COMERCIAL

☐ DECISÃO SINGULAR ☐ DECISÃO COLEGIADA

Nome(s) Empresarial(ais) igual(ais) ou semelhante(s):

☐ SIM ☐ SIM

_____ _____
_____ _____
_____ _____
_____ _____
_____ _____

☐ NÃO ___/___/_____ _____ ☐ NÃO ___/___/_____ _____
Data Responsável Data Responsável

Processo em Ordem
À decisão

___/___/_____
Data

_____
Responsável

DECISÃO SINGULAR

	2ª Exigência	3ª Exigência	4ª Exigência	5ª Exigência
☐ Processo em exigência. (Vide despacho em folha anexa)	☐	☐	☐	☐
☐ Processo deferido. Publique-se e arquive-se.				
☐ Processo indeferido. Publique-se.				

___/___/_____ _____
Data Responsável

DECISÃO COLEGIADA

	2ª Exigência	3ª Exigência	4ª Exigência	5ª Exigência
☐ Processo em exigência. (Vide despacho em folha anexa)	☐	☐	☐	☐
☐ Processo deferido. Publique-se e arquive-se.				
☐ Processo indeferido. Publique-se.				

___/___/_____ _____ _____ _____
Data Vogal Vogal Vogal

Presidente da _____ Turma

OBSERVAÇÕES

MARINELY DE PAULA BOMFIM
SECRETÁRIA GERAL

JUNTA COMERCIAL DO ESTADO DE MINAS GERAIS

Registro Digital

Capa de Processo

Identificação do Processo		
Número do Protocolo	Número do Processo Módulo Integrador	Data
20/305.240-4	MGN2066110248	27/05/2020

Identificação do(s) Assinante(s)	
CPF	Nome
818.097.436-72	ALBERTO HENRIQUE DUARTE COLARES
260.677.618-66	MAURICIO CAMPOS ZUARDI

MARINELY DE PAULA BOMFIM
SECRETÁRIA GERAL

KUNUMI SERVIÇOS EM TECNOLOGIA DA INFORMAÇÃO S.A.

CNPJ/MF Nº 24.477.718/0001-31

NIRE Nº 3130011433-3

("Companhia")

ATA DE REUNIÃO DO CONSELHO DE ADMINISTRAÇÃO
REALIZADA EM 13 DE ABRIL DE 2020

1. DATA, HORA E LOCAL: Aos 13 dias do mês de abril de 2020, às 09:00 horas, no endereço da Companhia, situado na Rua Professor José Vieira de Mendonça, nº 770, 2º andar, sala 208, Edifício Institucional do BH-TEC, Bairro Engenho Nogueira, na Cidade de Belo Horizonte, Estado de Minas Gerais, CEP 31.310-260.

2. CONVOCAÇÃO E PRESENÇAS: Dispensada a convocação tendo em vista a presença da totalidade dos membros do Conselho de Administração da Companhia, presencialmente via tele/videoconferência ou outro meio eletrônico ("Conselheiros").

3. MESA: Os trabalhos foram presididos pelo Sra Ana Paula Machado Pessoa e secretariados pelo, Sr. Ramon Dias de Azevedo.

4. ORDEM DO DIA: Deliberar sobre: (i) Destituição da diretoria com mandato em vigor (ii) Eleição dos membros para composição da diretoria conforme prazo de mandato definido no parágrafo segundo do artigo 14 do Estatuto Social (iii) a alteração do artigo 16 do Estatuto Social da Companhia e a respectiva consolidação da última versão do Estatuto Social da Companhia (iv) a prática de atos pela Diretoria.

5. INSTALAÇÃO E DELIBERAÇÃO: Instalada a reunião, após a ampla discussão das matérias, os membros do Conselho de Administração, por unanimidade de votos e sem quaisquer ressalvas ou restrições, decidiram:

(i) Aprovar a destituição dos atuais membros que compõe a diretoria da Companhia;

(ii) Eleger os novos membros para compor os cargos de (i) Diretor Presidente, (ii) Diretor Financeiro, (iii) Diretor de Tecnologia e (iv) Diretor(a) Acadêmico e Pesquisa para o mandato de 3 anos conforme previsto no artigo 14 parágrafo segundo do Estatuto Social da Companhia sendo eles respectivamente:

ALBERTO HENRIQUE DUARTE COLARES, brasileiro, designer, solteiro, portador da cédula de identidade MG 3617017, SSP/MG, portador do CPF/MF 818.097.436-72, com residência a Rua Santa Catarina 1340, ap.1301 – Lourdes, CEP;30170-081, Belo Horizonte/MG;

RAMOM DIAS DE AZEVEDO, brasileiro, casado, analista de sistemas, inscrito no CPF/MF sob o n. 013.932.366-00, portador da Carteira de Identidade n. MG-11586-050, expedida pela SSP/MG, residente e domiciliado na Rua Maria Heilbuth Surette, 370 – 802, Bairro Buritis, CEP 30.575-100, na cidade de Belo Horizonte/MG,

MAURICIO CAMPOS ZUARDI, brasileiro, designer, casado, inscrito nº do CPFMF sob o n. 260.677.618-66, portador da cédula de identidade 01379928203 DETRAN/MG, residente e domiciliado na Rua. Bergamota 470, ap 42C. Alto de Pinheiros, CEP 05468-915, São Paulo/SP., e

MARIA GABRIELLA GRABOWSKY SEILER, Brasileira, economista, solteira, inscrita no CPF/MF nº 091.215.247-85, portadora do documento de identidade RG 128130895 DETRAN/RJ, com domicílio/residência à Rua Irmão Gonçalo, 74 apto. 51 Jardim Das Bandeiras, CEP: 05439-080 São Paulo/SP

(iii) Aprovar a alteração da redação do artigo 16 do Estatuto Social da Companhia e a respectiva consolidação do referido documento;

(iv) Autorizar os diretores da Companhia a praticarem todos os atos que se fizerem necessários à fiel efetivação das matérias deliberadas nesta reunião.

6. DECLARAÇÃO DE DESIMPEDIMENTO. Os membros da Diretoria ora eleitos e empossados, conforme consta dos Termos de Posse lavrados no Livro de Atas de Reuniões da Diretoria, aceitaram o cargo e declararam, cada um deles, antecipadamente, sob as penas da lei, para fins do disposto nos parágrafos 1º e 4º do art. 147 da lei 6.404/76, e, nos incisos II do art. 37, da lei 8.934/94, cientes de que qualquer declaração falsa importa em responsabilidade criminal, que (1) não estão impedidos por lei especial, ou condenados por crime falimentar, de prevaricação, peita ou suborno, concussão, peculato, contra a economia popular, contra o sistema financeiro nacional, contra as normas de defesa da concorrência, contra as relações de consumo, a fé pública ou a propriedade, ou a pena ou condenação criminal que vede, ainda que temporariamente, o acesso a cargos públicos ou que os impeça de exercer atividades empresariais ou a administração de sociedade empresariais; (ii) possuir reputação ilibada; e (iii) não ocupam cargo em sociedade que possa ser considerada concorrente da Companhia, e não têm interesse conflitante com o da Companhia. Para os fins do artigo 149, § 2º, da Lei das Sociedades por Ações, declaram que receberão eventuais citações e intimações em processos administrativos e judiciais relativos a atos de sua gestão nos endereços indicados acima, sendo que eventual alteração será comunicada por escrito à Companhia.

7. ARQUIVAMENTO E PUBLICAÇÕES LEGAIS. Por fim, os Conselheiros deliberaram o arquivamento desta ata perante o Registro de Empresas e que as publicações legais sejam feitas e os livros societários transcritos.

8. ENCERRAMENTO: Nada mais havendo a tratar e inexistindo qualquer outra manifestação, foram encerrados os trabalhos e lavrada esta ata em forma de sumário, a qual, após lida e achada conforme, foi devidamente assinada em 04 (quatro) vias de igual teor e forma. Mesa: Sra. Ana Paula Machado Pessoa - presidente; Sr. Ramon Dias de Azevedo – secretário. Conselheiros Presentes: Ana Paula Machado Pessoa, Leonardo Pinheiro Gasparin, Fernando Alves de Oliveira, Ramon Dias de Azevedo, Nivio Ziviani, Alberto Henrique Duarte Colares.

Certifico que a presente é cópia fiel da ata original lavrada em livro próprio.

Belo Horizonte, 13 de abril de 2020.

Presidente da Mesa	**Secretário**
Ana Paula Machado Pessoa	Ramon Dias de Azevedo
Assinado por meio de certificado digital	*Assinado por meio de certificado digital*

JUNTA COMERCIAL DO ESTADO DE MINAS GERAIS

Registro Digital

Documento Principal

Identificação do Processo		
Número do Protocolo	Número do Processo Módulo Integrador	Data
20/305.240-4	MGN2066110248	27/05/2020

Identificação do(s) Assinante(s)	
CPF	Nome
865.873.407-25	ANA PAULA MACHADO PESSOA
013.932.366-00	RAMOM DIAS DE AZEVEDO

MARINELY DE PAULA BOMFIM
SECRETÁRIA GERAL

TERMO DE POSSE

Eu, **ALBERTO HENRIQUE DUARTE COLARES,** brasileiro, designer, solteiro, portador da cédula de identidade MG 3617017, SSP/MG, portador do CPF/MF 818.097.436-72, com residência a Rua Santa Catarina 1340, ap.1301 – Lourdes, CEP;30170-081, Belo Horizonte/MG, tendo sido eleito para o cargo de Diretor Presidente da **KUNUMI SERVIÇOS EM TECNOLOGIA DA INFORMAÇÃO S.A.**localizada na Cidade de Belo Horizonte, Estado de Belo Horizonte, na Rua Professor José Vieira de Mendonça, nº 770, 2º andar, sala 208, Edifício Institucional do BH-TEC, Bairro Engenho Nogueira, CEP 31.310-260, devidamente inscrita no CNPJ/MF sob nº 24.477.718/0001-31 ("Companhia"), na Reunião de Conselho de Administração realizada na presente data, para o mandato de 3 anos, declaro aceitar minha eleição e assumir o compromisso de cumprir fielmente todos os deveres inerentes ao meu cargo, de acordo com a lei e o Estatuto Social da Companhia, bem como declaro atender às disposições do artigo 147 da Lei nº 6.404/76, conforme alterada ("LSA"), pelo que firmo este Termo de Posse.

Declaro, outrossim, não estar incurso em nenhum dos crimes previstos em lei que me impeçam de exercer a atividade empresária, estando ciente do disposto no artigo 147 da LSA.

Para os fins do artigo 149, § 2º, da LSA, declaro que receberei eventuais citações e intimações em processos administrativos e judiciais relativos a atos de minha gestão no endereço acima indicado, sendo que eventual alteração será comunicada imediatamente por escrito à Companhia.

Belo Horizonte, 13 de abril de 2020.

_____

ALBERTO HENRIQUE DUARTE COLARES

Anexo

Identificação do Processo		
Número do Protocolo	Número do Processo Módulo Integrador	Data
20/305.240-4	MGN2066110248	27/05/2020

Identificação do(s) Assinante(s)	
CPF	Nome
818.097.436-72	ALBERTO HENRIQUE DUARTE COLARES

TERMO DE POSSE

Eu, **MARIA GABRIELLA GRABOWSKY SEILER**, brasileira, economista, solteira, inscrita no CPF/MF nº 091.215.247-85, portadora do documento de identidade RG 128130895 DETRAN/RJ, com domicílio/residência à Rua Irmão Gonçalo, 74 apto. 51 Jardim Das Bandeiras, CEP: 05439-080 São Paulo/SP, tendo sido eleita para o cargo de Diretora Acadêmica e Pesquisa da **KUNUMI SERVIÇOS EM TECNOLOGIA DA INFORMAÇÃO S.A.** localizada na Cidade de Belo Horizonte, Estado de Belo Horizonte, na Rua Professor José Vieira de Mendonça, nº 770, 2º andar, sala 208, Edifício Institucional do BH-TEC, Bairro Engenho Nogueira, CEP 31.310-260, devidamente inscrita no CNPJ/MF sob nº 24.477.718/0001-31 ("Companhia"), na Reunião de Conselho de Administração realizada na presente data, para o mandato de 3 anos, declaro aceitar minha eleição e assumir o compromisso de cumprir fielmente todos os deveres inerentes ao meu cargo, de acordo com a lei e o Estatuto Social da Companhia, bem como declaro atender às disposições do artigo 147 da Lei nº 6.404/76, conforme alterada ("LSA"), pelo que firmo este Termo de Posse.

Declaro, outrossim, não estar incursa em nenhum dos crimes previstos em lei que me impeçam de exercer a atividade empresária, estando ciente do disposto no artigo 147 da LSA.

Para os fins do artigo 149, § 2º, da LSA, declaro que receberei eventuais citações e intimações em processos administrativos e judiciais relativos a atos de minha gestão no endereço acima indicado, sendo que eventual alteração será comunicada imediatamente por escrito à Companhia.

Belo Horizonte, 13 de abril de 2020.

_____

MARIA GABRIELLA GRABOWSKY SEILER

MARINELY DE PAULA BOMFIM
SECRETÁRIA GERAL

Anexo

Identificação do Processo		
Número do Protocolo	Número do Processo Módulo Integrador	Data
20/305.240-4	MGN2066110248	27/05/2020

Identificação do(s) Assinante(s)	
CPF	Nome
091.215.247-85	MARIA GABRIELLA GRABOWSKY SEILER

MARINELY DE PAULA BOMFIM
SECRETÁRIA GERAL

TERMO DE POSSE

Eu, **MAURICIO CAMPOS ZUARDI**, brasileiro, designer, casado, inscrito nº do CPFMF sob o n. 260.677.618-66, portador da cédula de identidade 01379928203 DETRAN/MG, residente e domiciliado na Rua. Bergamota 470, ap 42C. Alto de Pinheiros, CEP 05468-915, São Paulo/SP, tendo sido eleito para o cargo de Diretor de Tecnologia da **KUNUMI SERVIÇOS EM TECNOLOGIA DA INFORMAÇÃO S.A.**localizada na Cidade de Belo Horizonte, Estado de Belo Horizonte, na Rua Professor José Vieira de Mendonça, nº 770, 2º andar, sala 208, Edifício Institucional do BH-TEC, Bairro Engenho Nogueira, CEP 31.310-260, devidamente inscrita no CNPJ/MF sob nº 24.477.718/0001-31 ("Companhia"), na Reunião de Conselho de Administração realizada na presente data, para o mandato de 3 anos, declaro aceitar minha eleição e assumir o compromisso de cumprir fielmente todos os deveres inerentes ao meu cargo, de acordo com a lei e o Estatuto Social da Companhia, bem como declaro atender às disposições do artigo 147 da Lei nº 6.404/76, conforme alterada ("LSA"), pelo que firmo este Termo de Posse.

Declaro, outrossim, não estar incurso em nenhum dos crimes previstos em lei que me impeçam de exercer a atividade empresária, estando ciente do disposto no artigo 147 da LSA.

Para os fins do artigo 149, § 2º, da LSA, declaro que receberei eventuais citações e intimações em processos administrativos e judiciais relativos a atos de minha gestão no endereço acima indicado, sendo que eventual alteração será comunicada imediatamente por escrito à Companhia.

Belo Horizonte, 13 de abril de 2020.

_____

MAURICIO CAMPOS ZUARDI

Anexo

Identificação do Processo		
Número do Protocolo	Número do Processo Módulo Integrador	Data
20/305.240-4	MGN2066110248	27/05/2020

Identificação do(s) Assinante(s)	
CPF	Nome
260.677.618-66	MAURICIO CAMPOS ZUARDI

MARINELY DE PAULA BOMFIM
SECRETÁRIA GERAL

TERMO DE POSSE

Eu, **RAMOM DIAS DE AZEVEDO**, brasileiro, casado, analista de sistemas, inscrito no CPF/MF sob o n. 013.932.366-00, portador da Carteira de Identidade n. MG-11586-050, expedida pela SSP/MG, residente e domiciliado na Rua Maria Heilbuth Surette, 370 – 802, Bairro Buritis, CEP 30.575-100, na cidade de Belo Horizonte/MG, tendo sido eleito para o cargo de Diretor Financeiro da **KUNUMI SERVIÇOS EM TECNOLOGIA DA INFORMAÇÃO S.A.**localizada na Cidade de Belo Horizonte, Estado de Belo Horizonte, na Rua Professor José Vieira de Mendonça, nº 770, 2º andar, sala 208, Edifício Institucional do BH-TEC, Bairro Engenho Nogueira, CEP 31.310-260, devidamente inscrita no CNPJ/MF sob nº 24.477.718/0001-31 ("Companhia"), na Reunião de Conselho de Administração realizada na presente data, para o mandato de 3 anos, declaro aceitar minha eleição e assumir o compromisso de cumprir fielmente todos os deveres inerentes ao meu cargo, de acordo com a lei e o Estatuto Social da Companhia, bem como declaro atender às disposições do artigo 147 da Lei nº 6.404/76, conforme alterada ("LSA"), pelo que firmo este Termo de Posse.

Declaro, outrossim, não estar incurso em nenhum dos crimes previstos em lei que me impeçam de exercer a atividade empresária, estando ciente do disposto no artigo 147 da LSA.

Para os fins do artigo 149, § 2º, da LSA, declaro que receberei eventuais citações e intimações em processos administrativos e judiciais relativos a atos de minha gestão no endereço acima indicado, sendo que eventual alteração será comunicada imediatamente por escrito à Companhia.

Belo Horizonte, 13 de abril de 2020.

_____

RAMOM DIAS DE AZEVEDO

Anexo

Identificação do Processo		
Número do Protocolo	Número do Processo Módulo Integrador	Data
20/305.240-4	MGN2066110248	27/05/2020

Identificação do(s) Assinante(s)	
CPF	Nome
013.932.366-00	RAMOM DIAS DE AZEVEDO

TERMO DE AUTENTICAÇÃO - REGISTRO DIGITAL

Certifico que o ato, assinado digitalmente, da empresa KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A, de NIRE 3130011433-3 e protocolado sob o número 20/305.240-4 em 28/05/2020, encontra-se registrado na Junta Comercial sob o número 7855420, em 29/05/2020. O ato foi deferido eletrônicamente pelo examinador Zulene figueiredo.

Certifica o registro, a Secretária-Geral, Marinely de Paula Bomfim. Para sua validação, deverá ser acessado o sitio eletrônico do Portal de Serviços / Validar Documentos (https://portalservicos.jucemg.mg.gov.br/Portal/pages/imagemProcesso/viaUnica.jsf) e informar o número de protocolo e chave de segurança.

Capa de Processo

Assinante(s)	
CPF	Nome
818.097.436-72	ALBERTO HENRIQUE DUARTE COLARES
260.677.618-66	MAURICIO CAMPOS ZUARDI

Documento Principal

Assinante(s)	
CPF	Nome
013.932.366-00	RAMOM DIAS DE AZEVEDO
865.873.407-25	ANA PAULA MACHADO PESSOA

Anexo

Assinante(s)	
CPF	Nome
818.097.436-72	ALBERTO HENRIQUE DUARTE COLARES

Anexo

Assinante(s)	
CPF	Nome
091.215.247-85	MARIA GABRIELLA GRABOWSKY SEILER

Anexo

Assinante(s)	
CPF	Nome
260.677.618-66	MAURICIO CAMPOS ZUARDI

A autencidade desse documento pode ser conferida no portal de serviços da jucemg informando o número do protocolo 20/305.240-4.

TERMO DE AUTENTICAÇÃO - REGISTRO DIGITAL

Anexo

Assinante(s)	
CPF	Nome
013.932.366-00	RAMOM DIAS DE AZEVEDO

Belo Horizonte. sexta-feira, 29 de maio de 2020

Documento assinado eletrônicamente por Zulene figueiredo, Servidor(a) Público(a), em 29/05/2020, às 12:26 conforme horário oficial de Brasília.

A autencidade desse documento pode ser conferida no portal de serviços da jucemg informando o número do protocolo 20/305.240-4.

O ato foi deferido e assinado digitalmente por :

Identificação do(s) Assinante(s)	
CPF	Nome
873.638.956-00	MARINELY DE PAULA BOMFIM

Belo Horizonte. sexta-feira, 29 de maio de 2020

MARINELY DE PAULA BOMFIM
SECRETÁRIA GERAL

“PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES E USO”

[01] Esta tecnologia refere-se a um processo de obtenção de modelos baseados em Aprendizado de Máquina que inclui meios para promover a diversificação das explicações do modelo. A tecnologia utiliza técnicas como a decomposição do espaço de dados em estruturas locais, do tipo “*backbone*”, formando um conjunto do tipo “*backbone features*”; representações de explicações de modelos (explicabilidade) utilizadas como critério de clusterização e seleção de protótipos dentro do espaço de modelos para promover a diversificação de explicações. Utiliza também combinação (“ensemble”) de modelos, dentre outras técnicas. As vantagens propiciadas pela tecnologia são principalmente: 1) agilidade na modelagem devido à criação de características relevantes ao problema, economizando tempo e recursos computacionais, favorecendo os ajustes de hiperparâmetros; 2) versatilidade, já que a metodologia pode ser utilizada em problemas de qualquer segmento e formato de banco de dados; 3) propicia o entendimento dos modelos devido à utilização de técnicas de explicabilidade; 4) viabiliza a diversificação das explicações do modelo. A tecnologia aplica-se no contexto de solução de problemas em que se recorre a modelos baseados em Aprendizado de Máquina, para aperfeiçoar a elaboração e o desempenho de tais modelos.

[02] A explicação de modelos visa permitir que o especialista seja capaz de julgar, não apenas as predições do modelo, mas principalmente as causas para tais predições, ou, mais precisamente, de que maneira as informações foram combinadas de forma a se chegar na predição. Por vezes, essa abordagem demonstrou situações nas quais o especialista foi capaz de julgar a pertinência das características e informações

usadas pelo modelo, e por vezes essa abordagem demonstrou situações nas quais os especialistas foram capazes de criar novas teorias, e dessa forma gerar valor a partir da tecnologia (Vanian, J. "Why Most Companies Are Failing at Artificial Intelligence: Eye on A.I.", Fortune, 2019: https://fortune.com/2019/10/15/why-most-companies-are-failing-at-artificial-intelligence-eye-on-a-i/; (Framingham, M. "IDC Survey Finds Artificial Intelligence to be a Priority for Organizations But Few Have Implemented an Enterprise-Wide Strategy", Business Wire, 2019: https://www.businesswire.com/news/home/20190708005039/en/).

[03] Há no estado da técnica algumas tecnologias dedicadas a possibilitar o entendimento dos modelos devido à utilização de técnicas de explicabilidade aplicadas no processo de elaboração de modelos baseados em inteligência artificial, como as apresentadas a seguir.

[04] O documento de patente US20210049503A1, intitulado "Meaningfully explaining black-box machine learning models", com data de prioridade de 13/08/2019, descreve uma metodologia para auxilio de elaboração de modelos de classificação baseados em Aprendizado de Máquina. A metodologia propõe o agrupamento de protótipos de modelos em "clusters" e organiza o modelo em "ensembles", cujos resultados de classificação são apresentados com os valores SHAP (“Shapley Additive Explanations”) para auxiliar na compreensão da contribuição de determinadas estruturas locais de dados no resultado de classificação. Porém, o agrupamento em "clusters" de US20210049503A1 segue um critério relacionado às variáveis de entrada do modelo após a etapa de aprendizado e obtenção de modelo.

[05] O documento de patente EP3918603A1, intitulado "Predicting blood metabolites", com data de prioridade de 31/01/2019, revela um método para predição da quantidade de um determinado metabólito no sangue de um paciente. A tecnologia utiliza como critério para formação

de “clusters” de dados (que denotam perfis metabolômicos identificados) a distância euclideana das correlações de Spearman. Dessa forma, EP3918603A1 não se aplica a “clusters” de protótipos de modelos. Os valores SHAP em EP3918603A1 são utilizados para selecionar as características ("features") que mais contribuíram na predição de um determinado modelo treinado.

[06] O documento de patente EP3803714A1, intitulado "Systems and methods for decomposition of non-differentiable and differentiable models", com data de prioridade de 08/06/2018, apresenta uma metodologia de explicação para modelos obtidos por técnicas de Aprendizado de Máquina. A tecnologia recorre ao uso de “ensembles” formadas por modelos heterogêneos (também identificados no texto de origem como sub-modelos) e as explicações são fornecidas por pontuações oriundas dos valores SHAP associados a um ou mais modelos formadores desta "ensemble". A “ensemble” é formada por um critério que utiliza um coeficiente de diferenciação entre modelos, mas este critério não foi tecnicamente detalhado no documento EP3803714A1.

[07] O documento de patente US2020097858, intitulado “Prediction explainer for ensemble learning”, de 22/09/2018, apresenta o uso de um método para explicar predição em um processo de predição de um “machine learning ensemble” que utiliza a decomposição do espaço de dados em múltiplos modelos treinados a partir das estruturas locais decompostas. O método também é capaz de avaliar e determinar a importância das “features” na predição do modelo.

[08] No estado da técnica não foram encontradas tecnologias com características semelhantes à tecnologia proposta neste pedido de patente, que apresenta representações vetoriais de explicações de modelos (explicabilidade), utilizadas para elaboração de meios para

promover diversidade de explicações, de modo a propiciar um maior entendimento dos modelos e suas predições.

[09] A tecnologia proposta tira proveito de uma característica até então não explorada que consiste no fato de que um dado fenômeno pode estar relacionado com subpopulações e estruturas locais de dados, dentro do espaço de dados. Tais estruturas e subpopulações podem indicar explicações e interpretações contrastantes e concorrentes do fenômeno estudado. Dessa forma, a criação de modelos baseados em Aprendizado de Máquina que viabilizem a diversificação das explicações como forma de contribuição complementar na decisão e previsão do modelo resultante pode elevar o desempenho do modelo. A presente tecnologia apresenta um meio técnico para realizar a diversificação das explicações do modelo.

BREVE DESCRIÇÃO DAS FIGURAS

[010] A **figura 1** apresenta uma visão geral das principais etapas do processo proposto. Os dados de entrada são apresentados na forma de uma matriz de dados tabular de ordem n x m, em que n representa as instâncias dos dados (Xn) e m representa as "features" (fn) associadas às instâncias; na etapa (a) realiza-se uma amostragem randômica de diversos conjuntos de “features” e treinam-se modelos a partir destes conjuntos de “features”; na etapa (b) cria-se um espaço de modelos H′ com os modelos criados na etapa “a”, na etapa (c) realiza-se o cômputo dos valores médios (w) SHAP para cada “feature” (f) indicando o impacto de cada “feature” na saída do modelo; na etapa (d) os modelos do espaço de modelos H′ são agrupados em “clusters” (aglomerados) utilizando-se como critério de agrupamento os valores médios SHAP de cada modelo e, em seguida, seleciona-se um modelo protótipo de cada “cluster” para formar uma combinação de modelos (“ensemble”) que

configurará como um modelo resultante, conforme explicitado na etapa (e).

[011] A **figura 2** apresenta uma representação do corpo humano, (a) parte anterior e (b) posterior, em que cada região é identificada por um número no intervalo de 1 a 53. No gráfico da figura (c) relacionam-se as regiões do corpo humano representadas por suas respectivas indicações numéricas anteriormente atribuídas com a frequência em que cada área foi reportada como alvo de dor. As frequências representadas no gráfico referem-se às populações A (barras verticais na cor púrpura) e B(traço na cor preta).

[012] A **figura 3** exibe uma representação do espaço de modelos H' pela técnica T-SNE (t – Distributed Stochastic Neighbor Embbeding). Cada ponto representa um modelo x' posicionado no gráfico conforme as probabilidades de desfechos em que haja um significante alívio de dor. Modelos com probabilidades semelhantes para os mesmos pacientes são agrupados próximos no espaço. As cores indicam os valores médios do parâmetro de avaliação da área sob a curva AUC (“Area Under Curve”) referente à validação cruzada (“cross-validation”). Pontos menores denotam menores valores de variância. Os espaços de modelos representados com a técnica T-SNE (identificados de A a F) foram obtidos utilizando os algoritmos de aprendizado “*XGboost*” (“Extreme Gradient Boosting”) nos diagramas A, C e E e “Random Forests” nos diagramas B, D e F utilizando-se “features” baseadas nas escalas “Global Impact Change” (GIC) (diagramas E e F), VAS 30 (diagramas A e B)/VAS 50 “Visual Analogue Scale” (VAS) (diagramas C e D) em que os números 30 e 50 representam uma redução mínima percentual na intensidade da dor de 30% e 50%, respectivamente.

[013] A **figura 4** exibe uma representação do espaço de modelos H' pela técnica T-SNE (t-Distributed stochastic Neighbor embbeding) após

agrupamento dos modelos em "clusters". Cada “cluster” é representado com uma cor distinta no espaço (H') de preferências. Os modelos foram obtidos com base na escala VAS 30. A formação dos "clusters" aconteceu de acordo com algoritmos de clusterização e critérios de formação dos “clusters” e os parâmetros configurados para um melhor valor de silhueta ("silhouette value"). Foi utilizado o algoritmo de aprendizado “XGboost” (“Extreme Gradient Boosting”). Os valores de silhueta obtidos para cada um dos critérios de clusterização (“Predictions” (nos diagramas A e B), “Features” (nos diagramas C e D) e “SHAP values” (nos diagramas E e F)) são: Predictions = (0,17; 0,01); Features = (0,05; 0,03); SHAP values = (0,83; 0,95), respectivamente, para os dois algoritmos de clusterização utilizados: Dendrograma e "DBScan clustering".

[014] A **figura 5** exibe uma representação do espaço de modelos H' pela técnica T-SNE (t-Distributed Stochastic Neighbor Embbeding) após agrupamento dos modelos em "clusters", cada “cluster” é representado com uma cor distinta no espaço (H') de preferências. Os modelos foram obtidos com base na escala VAS 30. A formação dos "clusters" aconteceu de acordo com algoritmos de clusterização e critérios de formação dos “clusters” e os parâmetros configurados para um melhor valor de silhueta ("silhouette value"). Foi utilizado o algoritmo de aprendizado “Random Forests”. Os valores de silhueta obtidos para cada um dos critérios de clusterização (“Predictions” (nos diagramas A e B), “Features” (nos diagramas C e D) e “SHAP values” (nos diagramas E e F)) são: Predictions = (0,17; 0,01); Features = (0,05; 0,03); SHAP values = (0,83; 0,95), respectivamente, para os dois algoritmos de clusterização utilizados: Dendrograma e "DBScan clustering".

[015] A **figura 6** apresenta dois gráficos de uma visualização do tipo "SHAP summary plots" associados a dois protótipos de modelos,

selecionados dentre um conjunto de oito protótipos e obtidos utilizando-se o algoritmo de aprendizado “XGboost”. Os rótulos (“labels”) dos modelos foram baseados na escala VAS 30. A escala de cores representa a magnitude do valor de cada "feature" e no eixo horizontal temos uma escala com os valores SHAP que denotam o impacto das "features" no resultado do modelo.

[016] A **figura 7** apresenta dois gráficos de uma visualização do tipo "SHAP decision plots" associados aos dois protótipos de modelos apresentados na figura 6. Os erros e acertos dos modelos foram expressos em uma escala de cores em que se tem nos extremos o seguinte: no lado esquerdo, na cor vermelha, estão representados os resultados do modelo que configuram como verdadeiros positivos e na cor azul os falsos negativos. No lado direito, na cor azul, estão representados os resultados do modelo que configuram como verdadeiros negativos e na cor vermelha os falsos positivos.

[017] A **figura 8** apresenta dois gráficos de uma visualização do tipo "SHAP summary plots" associados a dois protótipos de modelos (selecionados dentre um conjunto de oito protótipos) e obtidos utilizando-se o algoritmo de aprendizado “XGboost”. Os rótulos (labels) dos modelos foram baseados na escala VAS 50, apresentadas na coluna de rótulos à esquerda dos gráficos. A escala de cores representa a magnitude do valor de cada "feature" e no eixo horizontal temos uma escala com os valores SHAP que denotam o impacto das "features" no resultado do modelo.

[018] A **figura 9** apresenta dois gráficos de uma visualização do tipo "SHAP decision plots" associados aos dois protótipos de modelos apresentados na figura 8. Os erros e acertos dos modelos foram expressos em uma escala de cores em que se tem nos extremos o seguinte: No lado esquerdo, na cor vermelha, estão representados os

resultados do modelo que configuram como verdadeiros positivos e na cor azul os falsos negativos. No lado direito, na cor azul, estão representados os resultados do modelo que configuram como verdadeiros negativos e na cor vermelha os falsos positivos.

[019] A **figura 10** apresenta dois gráficos de uma visualização do tipo "SHAP summary plots" associados a dois protótipos de modelos (selecionados dentre um conjunto de oito protótipos) e obtidos utilizando-se o algoritmo de aprendizado “XGboost”. Os rótulos (labels) dos modelos foram baseados na escala GIC, apresentadas na coluna de rótulos à esquerda dos gráficos. A escala de cores representa a magnitude do valor de cada "feature" e no eixo horizontal temos uma escala com os valores SHAP que denotam o impacto das "features" no resultado do modelo.

[020] A **figura 11** apresenta dois gráficos de uma visualização do tipo "SHAP decision plots" associados aos dois protótipos de modelos apresentados na figura 10. Os erros e acertos dos modelos foram expressos em uma escala de cores em que se tem nos extremos o seguinte: No lado esquerdo, na cor vermelha, estão representados os resultados do modelo que configuram como verdadeiros positivos e na cor azul os falsos negativos. No lado direito, na cor azul, estão representados os resultados do modelo que configuram como verdadeiros negativos e na cor vermelha os falsos positivos.

[021] A **figura 12** apresenta um gráfico referente ao parâmetro de avaliação da área sob a curva AUC (“Area Under Curve”) delimitados pelas curvas nas cores verde, vermelha e preta que se referem às combinações de modelo (“ensembles”) seguintes: para o algoritmo de aprendizado e o critério: “Random Forests + SHAP”; para o algoritmo de aprendizado e o critério: “XBoost + SHAP” e para uma técnica de referência do estado da técnica identificada como “BENCH”

(“Biclustering-driven ENsemble of Classifiers”), respectivamente. No eixo das abscissas (x) estão representados os valores para o indicador falso positivo e no eixo das ordenadas (y) estão representados os valores para o indicador verdadeiro positivo.

DESCRIÇÃO DETALHADA DA TECNOLOGIA

[022] Esta tecnologia refere-se a um processo de obtenção de modelos baseados em Aprendizado de Máquina que inclui meios para promover a diversificação das explicações do modelo. A tecnologia utiliza técnicas como a decomposição do espaço de dados em estruturas locais, do tipo “*backbone*”, formando um conjunto do tipo “*backbone features*”; representações de explicações de modelos (explicabilidade) utilizadas como critério de clusterização e seleção de protótipos dentro do espaço de modelos para promover a diversificação de explicações. Utiliza também combinações (“ensemble”) de modelos, dentre outras técnicas. Apresenta-se adiante o processo e suas concretizações alternativas.

[023] Processo para elaboração de modelos de Aprendizado de Máquina baseado na diversificação das explicações formado pelas seguintes etapas:

a. Definir um espaço de dados como um conjunto de n pontos de dados da forma (x, y) n, tal que x ∈ Rd e fornece um vetor de característica {x1, x2, . . ., xd} e y é a saída para uma entrada x;

b. Decompor o espaço de dados em estruturas locais, do tipo “backbone”, de modo que haja um conjunto de características do tipo “backbone features”;

c. Obter modelos por meio de treinamento a partir dos dados que representam uma combinação de subespaços individualmente relacionados aos conjuntos de características definidos na etapa “b” e obter modelos por meio de processos de Aprendizado de Máquina a partir do espaço de dados descrito, garantindo-se a minimização de suas

respectivas funções objetivas f(x) de modo a propiciar a amostragem de modelos a partir do conjunto de modelos obtidos, utilizando como critério a minimização da função f(x′) tal que $x' \subseteq x$, $|x'| << |x|$, sendo que as características (“features”) que compõe cada modelo x′ são randomicamente selecionadas e submetidas a um processo de aprendizado que subsidia a criação de modelos (x′);

d. Realizar também um processo de amostragem dos modelos (x′) obtidos na etapa “c” de acordo com um critério baseado na medida de erro ℓ(x′) individual de cada modelo, procedendo-se a seleção pela comparação de um valor limiar de erro ϵ, utilizando-se o critério de inclusão $\ell(x') \leq \epsilon$, formando-se um espaço resultante de modelos (H′) contendo as possíveis explicações das predições relacionadas ao problema/fenômeno modelado;

e. Gerar, para cada modelo no espaço (H′), representações de suas preferências (p) contidas em um vetor n-dimensional $P(x') = \{p_1, p_2, \ldots, p_n\}$, em que p_i corresponde à probabilidade que o modelo x′ atribuiu ao ponto de dados i, de modo que os modelos em H′ sejam representativos das diversas estruturas locais existentes no espaço de dados, de modo complementar, onde a seleção dos modelos segundo o critério de desempenho $\ell(x') \leq \epsilon$ propiciará que a estrutura local correspondente seja devidamente explicada pelo modelo x′ correspondente;

f. Gerar, para cada modelo no espaço (H′), representações de suas explicações (e) contidas em um vetor d-dimensional $E(x') = \{e_1, e_2, \ldots, e_d\}$, onde e_i corresponde à influência que a característica xi exerce na predição realizada pelo modelo x′;

g. Agrupar os modelos do espaço (H′) em “clusters” (aglomerados) formados pelos modelos individuais contidos em H′ segundo um processo de formação dos aglomerados que obedece a um

critério de formação de tais aglomerados baseando-se na identificação de grupos de modelos que são internamente densos e também separados dos demais modelos em termos de seus fatores explicativos, isto é, dentro de cada “cluster” (aglomerado) os fatores explicativos são semelhantes, enquanto os fatores dentro de “clusters” disjuntos são diferentes;

h. Selecionar os modelos mais performantes (protótipos) dentro do conjunto de modelos que formam cada um dos “clusters” (aglomerado) utilizando como critério para seleção dos modelos a diversidade de explicações que pode ser denotada (métricas de diversidade) e implementada pela minimização do compartilhamento de características ("features") entre protótipos de modelos selecionados para compor a “ensemble”;

i. Formar uma combinação (“ensemble”) com os modelos (protótipos) selecionados na etapa “h” a partir de todos os “clusters” (aglomerados);

j. Atribuir a cada protótipo um voto ponderado que represente um artifício para validação de seu erro;

k. Considerar, como a predição da ”ensemble”, o rótulo (que está associado ao seu respectivo modelo) que obtiver a maior quantidade de votos.

[024] O processo para elaboração de modelos de Aprendizado de Máquina baseado na diversificação das explicações poderá utilizar em sua etapa “c” árvores com aumento de gradiente e a técnica “Random Forests” como algoritmos de aprendizagem.

[025] O processo para elaboração de modelos de Aprendizado de Máquina baseado na diversificação das explicações poderá utilizar em sua etapa “g” a maximização da distância euclidiana entre “clusters” em que se encontram os modelos dentro do espaço de modelos H′, inclusive

promovendo a maximização da distância euclidiana entre “clusters” por meio da escolha e definição de parâmetros concernentes ao processo de clusterização.

[026] O processo para elaboração de modelos de Aprendizado de Máquina baseado na diversificação das explicações poderá utilizar em sua etapa “g” como critério de clusterização de modelos os fatores de explicação denotados pelos valores SHAP (“SHapley Additive exPlanations”) ou outra métrica baseada nos valores SHAP dos modelos.

[027] O processo para elaboração de modelos de Aprendizado de Máquina baseado na diversificação das explicações poderá utilizar em sua etapa “g” valores SHAP (“SHapley Additive exPlanations”) para avaliar a importância das características (“features”) na predição do modelo.

[028] O processo para elaboração de modelos de Aprendizado de Máquina baseado na diversificação das explicações poderá utilizar em sua etapa “g” como critério de formação dos aglomerados a distância entre seus vetores de explicação e, para maximizar a coesão e separação dos “clusters”, utilizar um critério com base na distância envolvendo a preferência de modelo.

[029] O processo aqui definido poderá utilizar escalas de avaliação de dor como rótulos e base para definição de “features” na criação de modelos de Aprendizado de Máquina para predição acerca da evolução do alívio da dor.

Exemplo 1 – Tecnologia aplicada para predição da evolução do alívio da dor

[030] A dor nos faz perceber que algo está errado com nosso corpo. A Associação Internacional para Estudo da Dor IASP (“International Association for the Study of Pain”) define a dor como “uma experiência sensorial e emocional desagradável associada ao dano tecidual

potencial, ou descrita com base no dano". Se a dor durar além do tempo esperado para cura após cirurgia, trauma ou outra condição, então pode ser caracterizada como dor crônica. Não há paradigma universal aceito para prevenção e tratamento da dor crônica. A dor crônica é um problema de saúde pública que afeta 20 a 30% da população dos países ocidentais. A dor crônica geralmente classifica-se em uma das seguintes categorias: (1) neuropática, (2) nociceptiva ou (3) nociplásica.

[031] A dor neuropática ocorre quando há dano real ao nervo. Os nervos conectam a medula espinhal ao resto do corpo e permitem que o cérebro se comunique com a pele, os músculos e os órgãos internos. O desequilíbrio nutricional, o alcoolismo, presença de toxinas, infecções ou doenças autoimunes podem danificar essas conexões e causar dor. A dor neuropática também pode ser causada por um tumor cancerígeno pressionando um nervo ou um grupo de nervos. As pessoas muitas vezes descrevem essa dor como uma sensação de queimação ou peso, ou dormência ao longo do trajeto do nervo afetado. A dor nociceptiva é causada por danos nos tecidos do corpo e geralmente descrita como uma dor aguda, dolorida ou latejante. Este tipo de dor pode ser devido a patologia benigna, ou por tumores ou células cancerosas que estão crescendo e se aglomerando em outras partes do corpo que estão perto do local do câncer. A dor nociceptiva também pode ser causada por câncer se espalhando para os ossos, músculos ou articulações, ou que pode causar o bloqueio de um órgão ou vasos sanguíneos.

[032] A dor nociplásica surge da função nociceptiva alterada, apesar de não haver nenhuma evidência clara de dor real ou ameaça de dano tecidual causando a ativação de nociceptores periféricos ou evidência de lesão de doença do sistema somatossensorial que esteja causando a dor.

[033] Embora tenha havido muitos avanços científicos no entendimento da neurofisiologia da dor, definir com precisão a melhor terapia para um paciente ainda é um desafio. Atualmente, não existem estruturas clássicas validadas para o tratamento da dor. A dor crônica é uma experiência individualizada com etiologia multifatorial, e a compreensão dos contextos biológico, social, físico e psicológico é vital para o sucesso do tratamento. Instrumentos padronizados de autodeclaração e questionários para avaliar a intensidade da dor do paciente, habilidades funcionais, crenças, expectativas e sofrimento emocional estão disponíveis e podem ser usados para auxiliar no planejamento do tratamento.

[034] A dor é frequentemente avaliada em uma escala de “sem dor” a “pior dor imaginável”. A Escala Visual Análoga (“Visual Analogue Scale” - VAS), ou simplesmente EVA, é uma linha de 10 cm sem marcações que vai de “nenhuma dor” a “pior dor”. Os pacientes marcam sua pontuação de dor e uma medida em centímetros define seu nível de dor.

[035] No presente estudo, seiscentos e trinta e um participantes autopreencheram o questionário de dor de McGill “McGill Pain Questionnaire” (McGill) e a EVA. O questionário McGill avalia tanto a qualidade quanto a intensidade da dor. Em resumo, o questionário é composto por 78 palavras, das quais os respondentes escolhem aquelas que melhor descrevem a sua experiência de dor, sendo permitidas várias marcações (J. Hill, K. Dunn, M. Lewis, R. Mullis, C. Main, N. Foster, and E. Hay. 2008. A primary care back pain screening tool: Identifying patient subgroups for initial treatment. Arthr. Rheum 5, 59 (2008), 632–641.), (R. Melzack. 1975. The McGill pain questionnaire. Major properties and scoring methods. Pain 1 (1975), 277–299.). As palavras estão organizadas em três dimensões: (1) dimensão sensorial: engloba tanto a qualidade quanto a gravidade da dor em termos de suas propriedades

temporais, espaciais, de pressão e térmicas; (2) dimensão afetiva: refere-se aos sentimentos e sentimentos na presença da dor, ou seja, como o paciente se sente emocionalmente em decorrência da dor; (3) dimensão avaliativa: refere-se à avaliação global da situação vivenciada pelo paciente e é fortemente influenciada por experiências dolorosas anteriores. É uma avaliação subjetiva da intensidade geral da dor.

[036] Como resultado, os dados de dor incluem variáveis relacionadas à gravidade da dor, mudança no alívio da dor ao longo do tempo, radiação da dor, entre outras.

[037] Para o presente estudo foram também coletados, por meio de autodeclaração, dados sobre status socioeconômico (ou seja, idade, sexo), classificação global de saúde geral, fatores de risco conhecidos (ou seja, idade, tabagismo, consumo de álcool) e doenças concomitantes.

[038] Finalmente, os dados de dor coletados também incluíram as terapias prescritas pelo médico.

[039] Ao todo, os dados de dor do presente estudo são compostos por 338 variáveis sobre alívio da dor, status socioeconômico e tratamentos prescritos.

[040] Como etapa de pré-processamento, optamos por transformar “features” (características) categóricas em conjuntos de “features” do tipo “dummy” (por exemplo, presença ou ausência de um atributo). Acredita-se que este pré-processamento tem um impacto positivo na geração de modelos.

[041] No presente estudo, foram avaliadas três medidas distintas para identificar o sucesso no alívio da dor do paciente: (1) uma redução geral da intensidade da dor em 30% (também conhecida como VAS 30), que é formalmente considerada um resultado de tratamento bem-sucedido (R. Dworkin, D. Turk, K. Wyrwich, D. Beaton, D. Cleeland, J. Farrar, J.

Haythornthwaite, M. Jensen, R. Kerns, D. Ader, N. Brandenburg, L. Burke, D. Cella, J. Chandler, P. Cowan P., R. Dimitrova, R. Dionne, S. Hertz, A. Jadad, N. Katz, H. Kehlet, L. Kramer, D. Manning, C. McCormick, M. McDermott, H. McQuay, S. Patel, L. Porter, S. Quessy, B. Rappaport, C. Rauschkolb, D. Revicki, and M. Rothman, 2008. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J. Pain 9, 2 (2008), 105–121.), sendo que os rótulos de verdade são obtidos calculando a diferença das intensidades de dor relatadas na primeira e na última consulta; (2) uma redução geral da intensidade da dor em 50% (também conhecida como VAS 50), onde os rótulos de verdade são obtidos calculando a diferença das intensidades de dor relatadas na primeira e na última consulta; e (3) uma mudança do tipo mudança de impacto global (“Global Impact Change” - GIC) como uma escala de variação discreta de -3 a 3 fornecida pelo médico, indicando o grau de melhora no alívio da dor na opinião do médico, sendo que o sucesso é dado como um valor de pelo menos 2 na última consulta.

[042] No presente estudo calculou-se a correlação dois a dois entre GIC, VAS 30 e VAS 50. Observou-se que VAS 30 e VAS 50 são escalas altamente correlacionadas, alcançando um valor de correlação tão alto quanto 0,85. No entanto, o GIC não é altamente correlacionado com VAS 30 e VAS 50, apresentando valores de correlação de 0,1 e 0,097, respectivamente. Isso indica que as avaliações dos pacientes e a avaliação dos médicos podem ter discrepâncias. Além disso, quando um paciente atinge uma redução geral da intensidade da dor em 30%, na maioria das vezes também atingirá uma redução geral de 50%.

[043] É importante mencionar que, embora o tipo de dor tenha clara importância para o diagnóstico e tratamento, a resposta ao tratamento é mais uma característica pessoal dos pacientes e menos relacionada ao

tipo de dor em si. Em nosso estudo, os pacientes receberam tratamento de acordo com sua principal síndrome dolorosa, como parte de seus cuidados habituais, mas a etiologia da dor não foi um fator principal e não foi considerada diretamente em nossos modelos.

[044] Caracterização baseada na VAS 30: Como um resultado de tratamento bem-sucedido pode ser formalmente dado pela VAS 30, a análise apresentada ao longo desta seção refere-se especificamente ao rótulo (label) VAS 30. Considerando esse rótulo, dividimos os 631 pacientes do nosso estudo em duas populações: (1) população A: 277 pacientes para os quais o tratamento resultou em redução significativa no alívio da dor, ou seja, esses pacientes relataram uma redução significativa de +30% no alívio da dor após o término do tratamento; e (2) população B: 354 pacientes para os quais o tratamento não foi eficaz.

[045] A Tabela 1 mostra as características dos pacientes no conjunto de dados. A dor foi mais prevalente em mulheres e foi mais difícil obter redução significativa da dor em pacientes que relataram baixas intensidades iniciais de dor. A Tabela 1 também mostra as três dimensões da percepção da dor. As percepções de dor podem se sobrepor dentro da mesma dimensão, e uma pontuação total para cada dimensão é dada pela soma de todos os tipos de percepções de dor. Da mesma forma, a pontuação de McGill é dada como a soma dos valores associados a todas as palavras marcadas pelo paciente. A Tabela 1 também mostra a escala de dor neuropática que é usada para avaliar a dor neuropática e pode ser particularmente útil para avaliar a resposta a terapias. O escore total de neuropatia é calculado como a soma das possibilidades e o valor de corte para o diagnóstico de dor neuropática é uma pontuação total de 4. A Tabela 1 também mostra informações sobre os surtos de dor e o tempo em que a dor piora. Existem outras variáveis que foram omitidas da tabela para evitar

confusão. Na Tabela 1 o vocabulário das escalas VAS e McGill foram preservados na língua inglesa para não introduzir imprecisão ou ambiguidades, que são inevitáveis no processo de tradução, e poderiam prejudicar a compreensão da tecnologia. Uma tradução livre foi acrescentada para a língua portuguesa.

Tabela 1 - Dados dos pacientes obtidos na primeira consulta.

Escalas VAS e McGill	**Tradução livre das escalas VAS e McGill**	**Population A**	**Population B**
N	N	277 (43,89%)	354 (56,11%)
Sex (male)	Sexo (masculino)	110 (39,71%)	151 (42,65%)
Age, y	Idade, y	54,86 (46−64)	56,66 (45−60)
0−15 McGill score	0-15 pontuação McGill	7,21 (4−10)	5,75 (3−9)
0−10 intial pain intensity	0-10 intensidade inicial da dor	6,66 (5−8)	4,80 (2−8)
Sensory dimension	Dimensão sensorial	3,31 (1−5)	2,63 (1−4)
Burning	Queimando	170 (61,4%)	188 (53,1%)
Painful	Doloroso	131 (47,3%)	139 (39,3%)
Slapped	Pontadas	113 (40,8%)	115 (32,5%)
Throbbing	Latejante	111 (40,1%)	104 (29,4%)
Stabbings	Agulhadas	104 (37,5%)	96 (27,1%)
Electric	Choques	100 (36,1%)	99 (27,9%)

shocks	elétricos		
Sharp	Afiado	95 (34,3%)	102 (28,8%)
Spreads	Espalhado	87 (31,4%)	86 (24,3%)
Affective dimension	Dimensão afetiva	2.59 (1−3)	2.13 (1−3)
Tiring	Cansativo	209 (75,4%)	227 (64,1%)
Nauseous	Náusea	186 (67,1%)	191 (53,9%)
Annoying	Chato	157 (56,7%)	166 (46,9%)
Stifling	Sufocante	89 (32,1%)	91 (25,7%)
Scary	Apavorante	74 (26,7%)	79 (22,3%)
Evaluative dimension	Dimensão avaliativa	1,30 (1−2)	0,99 (1−1)
Uncomfortable	Desconfortável	260 (93,9%)	252 (71,2%)
Unbearable	Insuportável	100 (36,1%)	100 (28,2%)
Neuropathic pain scale	Escala de dor neuropática		
Burning	Queimando	193 (70,4%)	220 (62,7%)
Hypoesthesia to touch	Hipoestesia ao toque	143 (48,2%)	143 (40,7%)
Numbness	Dormência	109 (39,8%)	101 (28,8%)
Pins and needles	Comichão	107 (39,0%)	117 (33,3%)
Tingling	Formigamento	89 (32,5%)	97 (27,6%)
Electric shocks	Choques elétricos	85 (31,0%)	81 (23,1%)
Painful cold	Frio doloroso	46 (16,7%)	49 (14,0%)
Brushing	Coceira	40 (14,6%)	37 (10,5%)
Duration of	Duração dos		

pain outbreaks	surtos de dor		
Minutes	Minutos	10 (3,6%)	17 (4,8%)
Hours	Horas	19 (6,9%)	16 (4,5%)
Days	Dias	2 (0,8%)	5 (1,4%)
Weeks	Semanas	1 (0,4%)	3 (0,8%)
Months	Meses	6 (2,2%)	8 (2,3%)

[046] Por fim, a Figura 2 mostra a frequência com que a dor foi relatada em diferentes áreas do corpo humano. Curiosamente, as áreas do lado direito do corpo foram mais frequentemente relatadas pelos pacientes da população B. As características consideradas possibilitam inúmeras possibilidades de combinar diversos aspectos sobre o alívio da dor ao aprender modelos preditivos.

[047] Adiante discute-se o procedimento de avaliação e, em seguida, relatam-se os resultados. Em particular, os experimentos visam responder às seguintes questões de pesquisa: **RQ1**: Existe uma relação entre a explicação do modelo e as preferências do modelo?; **RQ2**: Os modelos de protótipos são diversos em termos de fatores explicativos?; **RQ3**: Pode-se construir “ensembles” (combinações de protótipos escolhidos com base em critérios definidos) eficazes combinando modelos que estão associados a diversos fatores explicativos?; **RQ4**: A abordagem de “ensemble” de diversificação de explicações é superior à abordagem de “ensemble” de “biclustering”?

[048] Configuração: Ao amostrar o espaço do modelo, define-se aleatoriamente o número de “features” que compõem cada modelo, mas garante-se que nenhum modelo tenha mais de 15 “features”. Há uma compensação ao ajustar o número máximo de “features”. A partir de experimentos, à medida que aumentamos o número máximo de

“features” permitidas, conseguimos obter “ensembles” com melhores valores de área sob a curva AUC (“Area Under Curve”). Por outro lado, espera-se também aumentar o custo computacional e dificultar a interpretação dos modelos gerados. Como a interpretabilidade é um aspecto crucial deste trabalho, opta-se por definir o limite superior em 15 “features” como um bom compromisso entre interpretabilidade e desempenho. Usando este limite superior, pode-se testar a validade e viabilidade do processo proposto. Além disso, pretende-se construir um questionário a ser aplicado pelo médico na primeira consulta do paciente e limitar o número de “features” também é uma forma de limitar o número de perguntas.

[049] As “features” que compõem cada modelo também foram selecionadas aleatoriamente. Os modelos foram construídos usando implementações SciKit-Learn dos algoritmos XGBoost ou “Random Forests” (F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R.Weiss, V. Dubourg, J. VanderPlas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12 (2011), 2825–2830.). Amostrou-se um total de 150.000 modelos usando o algoritmo XGBoost e outros 150.000 modelos usando o algoritmo “Random Forests”. Para avaliar o desempenho dos modelos, usou-se a medida padrão AUC ROC (Receiver Operating Characteristics) (T. Fawcett. 2006. An introduction to ROC analysis. Pattern Recog. Lett. 27, 8 (2006), 861–874.), (J. Hanley and B. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic ROC curve. Radiology 143 (1982), 29–36.). Realizou-se a validação cruzada de cinco vezes, ou seja, os dados foram organizados em cinco dobras (“folds”) e, em cada corrida, quatro dobras foram usadas como conjunto de treinamento e a dobra restante como conjunto de teste. Também

empregamos um conjunto de validação separado usado para selecionar os melhores modelos. Reporta-se o valor médio de AUC nas cinco corridas. Todo esse processo foi executado separadamente para cada rótulo, ou seja, VAS 30, VAS 50 e GIC.

[050] Modelos de referência (“baseline models”): Como referência, calcula-se a média dos valores de AUC pelos modelos baseados na estratégia “all-in-one” (“tudo em um só modelo”), que são modelos do estado da técnica que não se baseiam em “ensembles” ou combinações de protótipos de modelos obtidas com a estratégia de diversificação de explicações, como acontece na tecnologia ora proposta; e também utilizou-se a ferramenta de otimização “Treebased Pipeline Optimization Tool” (TPOT), (http://epistasislab.github.io/tpot/.). O primeiro cenário representa a abordagem padrão. O segundo cenário emprega uma ferramenta que otimiza o processo de Aprendizado de Máquina usando programação genética. Definiu-se o limite de tempo para otimização como 24 horas, uma vez que esse é o tempo aproximado para o pior caso.

[051] Fornecer como entrada a um modelo XGBoost todas as “features” e usar a VAS 30 como rótulo resultou em uma AUC média de 0,648. Com o algoritmo de aprendizado “Random Forests”, o valor obtido foi de 0,652. Portanto, considera-se um modelo VAS 30 com desempenho mínimo se seu valor médio de AUC for pelo menos igual ao valor de AUC da abordagem “all-in-one” (neste caso, 0,648 para o modelo XGBoost e 0,652 para “Random Forests”). Embora esse limite de desempenho pareça baixo, ele supera em muito o desempenho estimado do médico na primeira consulta, que não é superior a 0,584. O desempenho quase aleatório dos médicos na primeira consulta revela o quão difícil é essa tarefa preditiva.

Tabela 2 - Modelos utilizados como referência e seus valores AUC para os rótulos VAS30/50 e GIC.

Rótulo	XGBoost AUC	Random Forests AUC	Média	TPOT AUC
VAS 30	0,648	0,652	0,650	0,632
VAS 50	0,634	0,597	0,615	0,598
GIC	0,564	0,575	0,569	0,568

[052] Tabela 2 apresenta os valores médios de AUC obtidos usando a abordagem “all-in-one” para todos os rótulos. Vale ressaltar a dificuldade de prever o rótulo GIC. A ferramenta de otimização TPOT de processo de Aprendizado de Máquina seleciona automaticamente o algoritmo de aprendizado. A AUC média obtida usando VAS 30 como marcador foi de 0,632. Neste caso foi selecionado o classificador XGBoost. O uso do VAS 50 resultou em uma AUC média de 0,598, por meio da associação de vários estimadores: “Multinomial Naive Bayes”, “Gaussian Naive Bayes” e Classificador “k-Nearest Neighbors”. Finalmente, o uso de GIC como rótulo resultou em uma AUC média de 0,568 por meio da associação dos seguintes estimadores: “Stochastic Gradient Descent” (SGD) e XGBoost.

[053] O limite de desempenho resultou em um espaço (H') de modelo VAS 30 amostrado para XGBoost e outro espaço para “Random Forests”. O espaço do modelo XGBoost VAS 30 é composto por 2.830 modelos dos 150.000 modelos originais (1,9% dos modelos têm desempenho melhor do que o modelo “all-in-one”), enquanto o espaço do modelo “Random Forests” é composto por 2.507 modelos (1,7% dos modelos apresentam desempenho melhor do que o modelo “all-in-one”). Para o rótulo VAS 50, o número de modelos amostrados para o XGBoost

foi de 1.408 modelos (0,94% dos modelos têm desempenho melhor do que o modelo “all-in-one”) e para “Random Forests”, 11.829 (7,89% dos modelos têm desempenho melhor do que o modelo “all-in-one”). Em relação ao GIC, 18.575 modelos foram selecionados para XGBoost (12,38% dos modelos têm desempenho melhor do que os modelos “all-in-one”) e 10.035 modelos para “Random Forests” (6,69% dos modelos têm desempenho melhor do que o modelo “all-in-one”).

[054] A Figura 3 mostra os espaços de modelos XGBoost e “Random Forests” para cada rótulo, sendo que cada ponto corresponde a um modelo, e o tamanho do ponto indica a variância do erro de validação. Assim, nessa figura os melhores modelos são mostrados como pontos mais claros e menores. A figura mostra que os melhores modelos estão bem dispersos pelo espaço de modelos, indicando que existem modelos com preferências diferentes, mas com desempenhos iguais.

[055] Nos experimentos, observa-se que para que os algoritmos de agrupamento (clusterização) funcionem corretamente, são necessários pelo menos 1.000 elementos no espaço de explicação. Este número é obtido após filtrar este espaço com um limite AUC. Existem duas maneiras possíveis de alterar esse valor: (a) diminuir o limite usado para filtragem ou (b) aumentar o número máximo de “features”. A primeira alternativa deve ser aplicada com cautela, pois uma característica fundamental para que uma “ensemble” alcance um bom desempenho é que seus modelos base também tenham necessariamente um bom desempenho (Ludmila I. Kuncheva, Fabio Roli, Gian Luca Marcialis, and Catherine A. Shipp. 2001. Complexity of data subsets generated by the random subspace method: An experimental investigation. In International Workshop on Multiple Classifier Systems. Springer, 349–358.). A adição de modelos que não apresentam um bom desempenho pode ter um

impacto negativo no desempenho do conjunto. Portanto, é mais interessante aumentar o número máximo de “features”.

[056] Relacionando preferências de modelos e fatores explicativos: Para responder ao RQ1, incorporamos os modelos XGBoost e “Random Forests” de acordo com suas preferências de modelo (ou seja, probabilidades que eles atribuem aos pontos de dados), de modo que os modelos que atribuem probabilidades semelhantes aos mesmos pontos de dados são colocados próximos uns dos outros no espaço de modelos (como na Figura 3, mas usando o espaço original de alta dimensão). Em seguida, agrupamos os modelos do espaço de modelos usando diferentes critérios e algoritmos de agrupamento. Mais especificamente, foram empregados algoritmos de “Hierarchical clustering” (agrupamento hierárquico) e DBScan, (M. Ester, H. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In International Conference on Knowledge Discovery and Data Mining. 226–231.), (J. Ward. 1963. Hierarchical grouping to optimize an objective function. J. Amer. Statist. Assoc. 58 (1963), 236–244.). Esses dois algoritmos de agrupamento ou clusterização representam duas maneiras distintas de agrupar dados. “Dendrogram” (dendrograma) é um algoritmo de agrupamento baseado em árvore que particiona os dados fornecidos em vez de todo o espaço da instância. No entanto, o DBSCan é um agrupamento baseado em densidade que conecta pontos dentro de certos limites de distâncias, somente quando satisfaz um critério de densidade.

[057] Todos os hiperparâmetros foram definidos maximizando o valor da silhueta considerando o espaço de preferência do modelo. O valor da silhueta é uma medida de quão semelhante um ponto de dados é ao seu próprio “cluster” (coesão) em comparação com outros “clusters” (separação). A silhueta varia de -1 a +1, onde um valor alto indica que o

ponto de dados é bem compatível com seu próprio “cluster” e não compatível com “clusters” vizinhos. Se a maioria dos pontos de dados tiver um valor alto, a configuração de clusterização será apropriada. Essa medida foi escolhida por ser uma métrica padrão para detectar a má qualidade do agrupamento. O valor da silhueta considerado neste trabalho é o valor médio da silhueta sobre todas as amostras.

[058] A Figura 4 mostra a visualização 2D T-SNE para o espaço do modelo XGBoost VAS 30 após ser agrupado usando diferentes critérios. Os modelos foram agrupados no espaço de modelos baseando-se nas predições realizadas por cada modelo de modo que os modelos que realizam as mesmas predições para os mesmos pontos de dados tenham maior probabilidade de serem associados ao mesmo “cluster”. Neste caso, não foram utilizados fatores explicativos. Embora o agrupamento hierárquico leve à coesão, falta desempenho em termos de separação. A tendência oposta é observada para “clusters” DBScan. Tanto o agrupamento hierárquico quanto o DBScan alcançaram valores de silhueta baixos. Especificamente, o agrupamento hierárquico alcançou um valor de silhueta de 0,17, enquanto o DBScan conseguiu atingir apenas um valor de silhueta de 0,01. Os baixos valores de silhueta, especialmente para DBScan, podem ser devidos a probabilidades semelhantes que estão associadas a previsões opostas. Ou seja, pequenas diferenças entre probabilidades próximas ao limiar podem levar a previsões opostas, assim os modelos podem ser avaliados como semelhantes em termos de preferência de modelo, mas diferentes em termos de suas previsões.

[059] A Figura 4 também mostra o espaço do modelo XGBoost agrupado usando os índices das “features” dentro de cada modelo. Isso também é para fins de comparação, e a intuição é avaliar até que ponto um conjunto específico de “features” pode estar associado a uma estrutura

local específica no espaço de dados. O problema com este critério de agrupamento (ou seja, “features” dentro do modelo) é que ele negligencia que diferentes “features” podem ser correlacionados e, portanto, os modelos podem ter preferências semelhantes, mesmo que sejam completamente diferentes em termos de “features” (diagramas C e D). Novamente, isso leva a um desempenho pior do “cluster”. Especificamente, o agrupamento hierárquico (diagramas A, C e E) atingiu um valor de silhueta de 0,05, enquanto o DBScan (diagramas B, D e F) alcançou apenas um valor de silhueta de 0,03. Finalmente, avaliou-se a abordagem proposta de agrupar o espaço do modelo com base nos fatores explicativos associados a cada modelo (diagramas A e B). Representou-se cada modelo como um vetor composto pelos valores SHAP associados aos fatores que explicam as decisões do modelo. Curiosamente, o agrupamento baseado em fatores explicativos resulta em grupos com valores muito altos de coesão e separação, sugerindo uma forte ligação entre as preferências do modelo e a explicação do modelo. Outra possível vantagem de agrupar modelos como vetores de valores SHAP (diagramas F e E) é que a importância de cada fator é dividida se o modelo contiver características correlacionadas. Em particular, esta abordagem evita uma instabilidade sistemática em que modelos semelhantes em termos de preferências podem ter explicações muito diferentes. Como consequência, os valores de silhueta são tão altos quanto 0,83 para agrupamento hierárquico e 0,95 para DBScan. De fato, a figura mostra pequenas diferenças na configuração dos grupos obtidos por ambos os algoritmos de agrupamento. Como resultado, a resposta ao RQ1 é positiva considerando o espaço do modelo XGBoost, pois os altos valores de silhueta para ambos os algoritmos de agrupamento indicam uma relação entre as preferências do modelo e a explicação do modelo.

[060] Da mesma forma, a Figura 5 mostra a visualização 2D T-SNE para o espaço do modelo “Random Forests” após ser agrupado usando diferentes critérios. A mesma tendência foi observada. Novamente, para fins de comparação, agrupamos o espaço do modelo usando as previsões realizadas por cada modelo. Neste caso, novamente, tanto o agrupamento hierárquico quanto o DBScan alcançaram valores de silhueta baixos. Especificamente, o agrupamento hierárquico atingiu um valor de silhueta de 0,06, enquanto o DBScan atingiu um valor de -0,33 de silhueta. Agrupar o espaço do modelo usando a distância entre os conjuntos de “features” dentro de cada modelo também leva a uma coesão e separação ruins. Nesse caso, o agrupamento hierárquico atingiu um valor de silhueta de 0,04, enquanto o DBScan atingiu um valor de silhueta de 0,06. Novamente, o agrupamento baseado em fatores explicativos resulta em grupos com valores muito altos de coesão e separação. Os valores de silhueta são tão altos quanto 0,87 para “cluster” hierárquico e 0,98 para DBScan. Portanto, a resposta ao RQ1 também é positiva considerando o espaço do modelo “Random Forests”, pois os altos valores de silhueta para ambos os algoritmos de agrupamento indicam uma relação entre as preferências do modelo e a explicação do modelo.

[061] Resultados semelhantes foram obtidos considerando também os marcadores VAS 50 e GIC. Especificamente para o VAS 50 com o modelo XGBoost, a pontuação da silhueta ao agrupar com critério de recurso é 0,017 para agrupamento hierárquico e -0,021 para DBScan. Para o critério de probabilidade, os valores de 0,089 e -0,288, respectivamente, são obtidos com agrupamento hierárquico e DBScan. Por fim, considerando o critério SHAP, obtêm-se bons valores de silhueta de 0,8115 e 0,95, mostrando novamente a coesão e separação obtidas quando usamos fatores explicativos como critério. Quando utilizamos o

modelo “Random Forests”, os valores das silhuetas seguem o mesmo padrão. Para o critério de características, obtemos 0,0055 e −0,0076, para o critério de probabilidade 0,0377 e −0,3741; por último, o critério do fator explicativo, obtemos 0,7839 e 0,8895. Sempre o primeiro valor referente ao algoritmo de agrupamento hierárquico e o segundo ao algoritmo de agrupamento DBScan.

[062] Para GIC com modelo XGBoost, a pontuação de silhueta ao agrupar com critério baseado em “features” é 0,0055 para agrupamento hierárquico e -0,0175 para DBScan. Para o critério de probabilidade os valores de 0,2311 e -0,4190, respectivamente, são obtidos com agrupamento hierárquico e DBScan. Por fim, considerando o critério SHAP, obtêm-se valores de silhueta de 0,3762 e 0,0064. Quando utilizamos o modelo “Random Forests”, os valores das silhuetas seguem o mesmo padrão. Para o critério de características, obtemos 0,01779 e 0, -0041, para o critério de probabilidade 0,1592 e 0,2454; por último, o critério do fator explicativo, obtemos 0,4027 e 0,2167.

[063] Embora os valores de silhueta ao agrupar pelo critério do fator explicativo sejam no máximo 0,4027 (menores quando comparados aos rótulos VAS 30 e VAS 50), é importante notar que este superou outros critérios para GIC. O menor valor de silhueta quando comparado aos demais rótulos provavelmente se deve à maior dificuldade de previsão desse rótulo. No entanto, notadamente, “clusters” com fatores explicativos mostram-se consistentemente superiores ao critério de baseado em “features” e probabilidades, independentemente do modelo e algoritmo de clusterização utilizado. Assim, podemos concluir que RQ1 também é positivo para os rótulos alternativos VAS 50 e GIC.

[064] Estrutura “backbone”, fatores explicativos e diversidade: Para responder à RQ2, inspecionou-se os modelos protótipos dentro de cada “cluster” nos espaços de modelo XGBoost e “Random Forests”. Os

“clusters” baseados em vetores de explicação produzidos pelo DBScan são os que levarão aos melhores resultados da “ensemble”. A Figura 6 mostra dois gráficos “SHAP summary plots” representativos associados a modelos de protótipo gerados no espaço do modelo VAS 30, fornecendo uma visão geral de quais “features” são mais importantes para um modelo (Esses gráficos mostram os valores SHAP de cada “feature” para cada ponto de dados. Em cada gráfico, os “features” são classificados pela soma das magnitudes dos valores SHAP em todos os pontos de dados. A cor representa o valor da “feature”.). Mostra-se apenas gráficos de resumo para modelos XGBoost para evitar confusão.

[065] O primeiro modelo da Figura 6 mostra que uma alta intensidade inicial de dor aumenta as chances de redução significativa da dor ao final do tratamento. Normalmente, a “feature” mais importante dentro de um modelo é uma “feature” do tipo backbone e, em seguida, o modelo inclui “features” que estão de alguma forma relacionados a uma “feature” do tipo backbone, como um local específico ou um medicamento específico. Como resultado, os modelos diferem muito em termos de seus fatores explicativos. A diversidade fica clara à medida que inspecionamos os modelos de protótipos, pois cada modelo emprega um conjunto de “features” muito diferente dos “features” usados pelos outros modelos de protótipos. Especificamente, dentro dos oito modelos de protótipos XGBoost, há um total de 42 “features” distintos, e apenas 5 “features” estão presentes em dois modelos. Assim, 37 características ocorrem apenas em um modelo protótipo. Também foram inspecionados os modelos protótipos dentro de cada “cluster” no espaço do modelo “Random Forests” VAS 30 (embora não sejam mostrados os gráficos de resumo correspondentes). Assim como no XGBoost, foca-se em “clusters” baseados em vetores de explicação produzidos pelo DBScan. Novamente, os modelos diferem muito, dependendo da dimensão da dor,

da localização da dor e sua predominância. A diversidade também é observada nesses modelos de protótipos. Especificamente, há um total de 46 “features” distintos dentro dos 10 modelos protótipos, dos quais 7 “features” estão presentes em dois modelos e apenas 2 “features” estão presentes em três modelos.

[066] A Figura 7 mostra os gráficos (SHAP Decisions) dos modelos de protótipo mostrados anteriormente na Figura 6. Os gráficos de decisão SHAP mostram como os modelos complexos chegam às suas previsões. Cada linha apresenta um caminho de decisão do modelo dada uma entrada. Cada gráfico de decisão foi gerado para 50 entradas, para melhor representar o comportamento do modelo. Essa visão diferente é efetiva principalmente quando muitas “features” significativos estão envolvidas. Como pode ser visto, os modelos que usam “features” diferentes representam formas totalmente diferentes de alcançar a previsão.

[067] Novamente, observa-se a mesma tendência para os modelos de protótipos obtidos para os rótulos (labels) VAS 50 e GIC. A Figura 8 mostra gráficos de resumo para dois dos nove modelos protótipos XGBoost usando o rótulo VAS 50. Neste caso, os protótipos XGBoost compreendem um total de 91 “features”, das quais 76 são “features” únicas.

[068] Para “Random Forests”, o número total de “features” dentro dos modelos correspondentes é de 155, dos quais 123 são únicas e ocorreram em apenas um modelo. Novamente, esta é uma forte indicação de que quando agrupados por fatores explicativos, cada protótipo que representa um “cluster” é diverso, uma estratégia crucial para construir modelos de “ensembles” eficazes. A Figura 9 apresenta os gráficos de decisão SHAP para os dois modelos de representantes usando o rótulo VAS 50.

[069] A Figura 10 mostra gráficos de resumo para os dois modelos protótipos XGBoost usando o rótulo GIC. Nesse caso, os protótipos do XGBoost compreendem um total de 13 “features” divididas em apenas dois protótipos. Curiosamente, para “Random Forests”, foram gerados 19 protótipos, resultando em 203 “features” utilizadas, das quais 143 eram exclusivas. Novamente, nossa estratégia de “ensemble” mostra empregar informações diversas ao construir o modelo final. A Figura 11 mostra os gráficos de decisão SHAP para os dois modelos de representantes usando o rótulo GIC. Nossa resposta ao RQ2 é positiva considerando os espaços de modelo XGBoost e “Random Forests”, pois os modelos de protótipo se diferem muito em termos das “features” usadas.

[070] Como os modelos foram selecionados maximizando a diversidade de explicações, espera-se que o número de “features” compartilhadas entre eles seja pequeno. As “features” mais relevantes para cada protótipo podem ser obtidas diretamente de seus valores de SHAP (ou seja, quanto maior o valor de SHAP, mais importante é a “features”). O conjunto de “features” mais relevante dentro do modelo final seria a combinação das “features” mais relevantes dentro de seus modelos de protótipo. Por exemplo, usando VAS30 como rótulo, XGBoost como algoritmo de aprendizado e DBScan como algoritmo de agrupamento (clusterização), o conjunto formado pela seleção da “feature” mais relevante de cada modelo protótipo é formado por: intensidade da dor, dimensão avaliativa, dimensão afetiva, DN4 quantitativa, dimensão avaliativa desconfortável, dimensão McGill e dimensão sensível.

[071] Desempenho da “ensemble”: O próximo conjunto de experimentos é dedicado a responder RQ3. As Tabelas 3, 4 e 5 mostram os valores de AUC para diferentes configurações de “ensembles”. O desempenho da “ensemble” é comparado com o desempenho do melhor modelo local no

espaço de modelos correspondente. Considerando o rótulo VAS 30, diferentes “ensembles” alcançaram valores de AUC que variam de 0,68 a 0,78. As “ensembles” obtidas a partir de modelos protótipos “Random Forests” proporcionaram ganhos de até 4,17%, embora para algumas configurações de “ensembles” o desempenho tenha se deteriorado. As “ensembles” obtidas a partir de modelos protótipos XGBoost foram mais eficazes, proporcionando ganhos de até 9,86% em relação ao melhor modelo no espaço do modelos. Se compararmos as “ensembles” com os modelos “all-in-one”, os ganhos são de até 21%, uma grande melhoria se considerarmos que os modelos também são mais simples. Para o rótulo VAS 30, os valores de AUC considerados como referência para XGBoost foi 0,71 e para “Random Forest”s 0,72. Para o rótulo VAS 50, os valores de AUC considerados como referência para XGBoost foi 0,68 e para “Random Forests” 0,71. Para o rótulo GIC, os valores de AUC considerados como referência para XGBoost foi 0,67 e para “Random Forests” 0,68.

Tabela 3 - Performance das "ensembles" para diferentes critérios e algoritmos de clusterização utilizando o rótulo VAS 30.

			XGBoost			Random Forests	
Critério	Clusterização	AUC	Ganhos (Maior)	Ganhos all-in-one	AUC	Ganhos (Maior)	Ganhos all-in-one
Predições	DBScan	0,73	2,82%	12,65%	0,68	-5,55%	4,29%
Predições	Hierarchical	0,73	2,82%	12,65%	0,73	1,39%	11,96%
Feature (valor)	DBScan	0,71	-	9,57%	0,70	-2,78%	7,36%
Feature (valor)	Hierarchical	0,72	1,51%	11,11%	0,74	2,78%	13,50%
Explicações	DBScan	0,78	9,86%	20,37%	0,75	4,17%	15,03%
Explicações	Hierarchical	0,77	8,45%	18,83%	0,75	4,17%	15,03%

Tabela 4 - Performance das "ensembles" para diferentes critérios e algoritmos de clusterização utilizando o rótulo VAS 50.

		XGBoost			Random Forests		
Critério	Clusterização	AUC	Ganhos (Maior)	Ganhos all-in-one	AUC	Ganhos (Maior)	Ganhos all-in-one
Predições	DBScan	0,69	0,73%	8,83%	0,66	-7,04%	10,55%
Predições	Hierarchical	0,79	15,33%	24,60%	0,72	1,41%	20,60%
Feature (valor)	DBScan	0,73	6,57%	15,14%	0,69	-2,82%	15,58%
Feature (valor)	Hierarchical	0,77	12,41%	21,45%	0,79	11,26%	32,33%
Explicações	DBScan	0,77	12,41%	21,45%	0,79	11,26%	32,33%
Explicações	Hierarchical	0,77	12,41%	21,45%	0,78	9,86%	30,65%

Tabela 5 - Performance das "ensembles" para diferentes critérios e algoritmos de clusterização utilizando o rótulo GIC.

		XGBoost			Random Forests		
Critério	Clusterização	AUC	Ganhos (Maior)	Ganhos all-in-one	AUC	Ganhos (Maior)	Ganhos all-in-one
Predições	DBScan	0,70	4,48%	24,11%	0,66	-2,97%	14,78%
Predições	Hierarchical	0,70	4,48%	24,11%	0,72	5,88%	25,22%
Feature (valor)	DBScan	0,65	-2,98%	15,24%	0,67	-1,47%	16,52%
Feature (valor)	Hierarchical	0,68	1,49%	20,57%	0,69	1,47%	20,00%
Explicações	DBScan	0,68	1,49%	20,57%	0,76	11,76%	32,17%
Explicações	Hierarchical	0,71	5,97%	25,89%	0,74	8,82%	28,70%

[072] Para os rótulos VAS 50 e GIC, as “ensembles” obtidas tiveram um desempenho melhor do que o melhor modelo dentro do espaço de modelo correspondente. Embora, por exemplo, usando o critério de probabilidade e algoritmo de agrupamento hierárquico, tenha sido possível obter um ganho de até 12,40% para o rótulo VAS 50 usando

XGBoost (superior ao critério de fatores explicativos), o resultado não foi consistente. Se apenas mudarmos o modelo para “Random Forests”, em vez de um grande ganho, teremos uma perda considerável de 2,81%. No entanto, as “ensembles” obtidas por agrupamento de acordo com os fatores explicativos do modelo em todos os casos resultaram em ganhos significativos, mesmo alterando o rótulo, o modelo ou o algoritmo de agrupamento. Um desempenho de melhoria de até 32% foi alcançado em comparação com os modelos “all-in-one”.

[073] A estratégia proposta de treinar modelos e gerar “ensembles” agrupando o espaço dos modelos usando fatores explicativos se mostrou eficaz, proporcionando ganhos significativos, seja qual for a configuração da “ensemble”. Assim, nossa resposta ao RQ3 é definitivamente positiva.

[074] Comparação com Desempenho Médico e a abordagem Biclustering: O último conjunto de experimentos é dedicado a responder RQ4. Para isso, considera-se a atuação do médico como sendo o resultado conhecido da última consulta. Além disso, como referência principal, consideramos o método BENCH (Biclustering-driven ENsemble of Classifiers) proposto na Referência (T. Pansombut, W. Hendrix, Z. Jacob Gao, B. Harrison, and N. Samatova. 2011. Biclustering-driven ensemble of Bayesian belief network classifiers for underdetermined problems. In International Joint Conference on Artificial Intelligence. 1439–1445.), que constrói um conjunto de classificadores por meio de recursos concorrentes e seleção de pontos de dados guiados por biclustering. A Figura 12 mostra as curvas ROC para BENCH, XGBoost + Explicações (ou seja, conjunto de modelos XGBoost agrupados por fatores de explicação SHAP) com DBScan e “Random Forests” + Explicações com DBScan. Os conjuntos de diversificação de explicação superam o BENCH em todas as faixas de taxas de falso positivo e verdadeiro positivo. Realizamos os testes t de Welch com p = 0,01, e

ambas as configurações de “ensemble” são estatisticamente diferentes de BENCH e, portanto, nossa resposta ao RQ4 também é positiva.

[075] Limitações: A limitação atual na estrutura ocorre principalmente no cálculo dos valores de Shapley. Os dois algoritmos de aprendizado baseados em árvore usados neste trabalho (“Random Forests” e “Gradient Boosted Trees”) dependem do TreeSHAP. Embora o KernelSHAP seja um método agnóstico de modelo para calcular os valores de Shapley, ele pode ser lento e sofrer com a variabilidade de amostragem. TreeSHAP, focando especificamente em árvores, computa explicações locais baseadas em valores exatos de Shapley em tempo polinomial (Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 1 (Jan. 2020), 56–67. DOI: https://doi.org/10.1038/s42256-019-0138-9). O uso de modelos que são dependentes do KernelSHAP para computar explicações locais atualmente não é viável. O KernelSHAP é duas ordens de magnitude mais lento que o TreeSHAP, inviabilizando seu uso no framework. Além disso, os custos computacionais também são inerentes ao framework, pois são gerados aproximadamente 150.000 modelos no total. Isso requer que o algoritmo de aprendizado escolhido seja capaz de gerar um modelo treinado em um período de tempo muito curto.

[076] Uma Nota sobre Requisitos de Tempo: Embora o tempo gasto para gerar as “ensembles” não influencie nos benefícios já demonstrados ao longo do trabalho, como aumento da AUC e redução do conjunto de “features”, pode ser um fator crítico na viabilidade de aplicação da técnica em casos de uso diversos. Quase toda a contribuição do tempo de execução vem da geração do espaço do modelo com 150.000

modelos amostrados. À medida que aumentamos o número máximo de “features” permitidas, conseguimos obter “ensembles” com desempenho aprimorado. Por outro lado, espera-se também aumentar o custo computacional. Como a interpretabilidade é um aspecto crucial do nosso trabalho, optamos por definir o limite superior para 15 “features”. O tempo total gasto para amostrar todo o espaço do modelo com limite superior de 15 recursos foi de 1.353,78 minutos com XGBoost e apenas 249,01 minutos com “Random Forests”. Ambos os casos usando VAS 30 como rótulo. A geração da “ensemble” a partir do espaço do modelos amostrado leva menos de 60 minutos para todas as configurações e algoritmos de aprendizado. As especificações relacionadas ao hardware em que foram executados os trabalhos são: Intel® Core™ i3-6100 CPU @ 3,70 GHz, 16 GB DDR3 1.600 MT/s e 256 GB SSD. Como os algoritmos utilizados não faziam uso da placa gráfica, omitiremos as informações.

[077] Conclusões: A tecnologia explora uma ligação pouco estudada entre modelagem explicativa e modelagem preditiva, o que leva a uma nova abordagem para aprendizagem em “ensembles”. A abordagem proposta explora dois conceitos: (i) os modelos locais que compõem a “ensemble” devem ser diversos em termos de seus fatores explicativos, e (ii) os modelos candidatos devem ser organizados buscando estabilidade no sentido de que os modelos que realizam previsões semelhantes também devem ser semelhantes em termos de seus fatores explicativos. Outra importante contribuição da tecnologia é a avaliação de nossa abordagem de aprendizagem em “ensemble” na tarefa de prever a evolução do alívio da dor em pacientes com condições de dor crônica desconhecida. Este é um problema em que normalmente existe um conjunto específico de “features” do tipo “backbone” que, uma vez definidas, fazem com que o restante das “features” se decomponha em

diferentes subconjuntos no espaço de dados. A estrutura do “backbone” sugere que o problema é definido por múltiplas “features” locais, sendo assim um exemplo motivador para a nossa abordagem de “ensemble learning”, que relaciona “features” locais e explicações de modelos. Os experimentos revelaram que a abordagem de “ensemble” proposta fornece ganhos de desempenho de até 11% quando comparados com os melhores modelos locais, e também supera significativamente uma abordagem de “biclustering”, proporcionando ganhos de 6,8%. Ao comparar com abordagens “all-in-one”, os ganhos são de até 32%.

REIVINDICAÇÕES

1. PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES, caracterizado por compreender as seguintes etapas:

a. Definir um espaço de dados como um conjunto de n pontos de dados da forma (x, y) n, tal que $x \in R^d$ e fornece um vetor de característica $\{x_1, x_2, \ldots, x_d\}$ e y é a saída para uma entrada x;

b. Decompor o espaço de dados em estruturas locais, do tipo "*backbone*", de modo que haja um conjunto de características do tipo "*backbone features*";

c. Obter modelos por meio de treinamento a partir dos dados que representam uma combinação de subespaços individualmente relacionados aos conjuntos de características definidos na etapa "b" e obter modelos por meio de processos de Aprendizado de Máquina a partir do espaço de dados descrito, garantindo-se a minimização de suas respectivas funções objetivas f(x) de modo a propiciar a amostragem de modelos a partir do conjunto de modelos obtidos, utilizando como critério a minimização da função f(x′) tal que $x' \subseteq x$, $|x'| << |x|$, sendo que as características ("features") que compõe cada modelo x′ são randomicamente selecionadas e submetidas a um processo de aprendizado que subsidia a criação de modelos (x′);

d. Realizar também um processo de amostragem dos modelos (x′) obtidos na etapa "c" de acordo com um critério baseado na medida de erro ℓ(x′) individual de cada modelo, procedendo-se a seleção pela comparação de um valor limiar de erro ϵ, utilizando-se o critério de inclusão $\ell(x') \leq \epsilon$, formando-se um espaço resultante de modelos (н′) contendo as possíveis explicações das predições relacionadas ao problema/fenômeno modelado;

e. Gerar, para cada modelo no espaço (н′), representações de suas preferências (p) contidas em um vetor n-dimensional P (x′) = {p_1, p_2, . . ., p_n}, em que p_i corresponde à probabilidade que o modelo x′ atribuiu ao ponto de dados i, de modo que os modelos em H′ sejam representativos das diversas estruturas locais existentes no espaço de dados, de modo complementar, onde a seleção dos modelos segundo o critério de desempenho ℓ (x′) ≤ ϵ propiciará que a estrutura local correspondente seja devidamente explicada pelo modelo x′ correspondente;

f. Gerar, para cada modelo no espaço (н′), representações de suas explicações (e) contidas em um vetor d-dimensional E (x′) = {e_1, e_2, . . ., e_d}, onde e_i corresponde à influência que a característica x_i exerce na predição realizada pelo modelo x′;

g. Agrupar os modelos do espaço (н′) em “clusters” (aglomerados) formados pelos modelos individuais contidos em н′ segundo um processo de formação dos aglomerados que obedece a um critério de formação de tais aglomerados baseando-se na identificação de grupos de modelos que são internamente densos e também separados dos demais modelos em termos de seus fatores explicativos, isto é, dentro de cada “cluster” (aglomerado) os fatores explicativos são semelhantes, enquanto os fatores dentro de “clusters” disjuntos são diferentes;

h. Selecionar os modelos mais performantes (protótipos) dentro do conjunto de modelos que formam cada um dos “clusters” (aglomerados), utilizando como critério para seleção dos modelos a diversidade de explicações que pode ser denotada (métricas de diversidade) e implementada pela minimização do compartilhamento de características ("features") entre protótipos de modelos selecionados para compor a “ensemble”;

i. Formar uma combinação (“ensemble”) com os modelos (protótipos) selecionados na etapa “h” a partir de todos os “clusters” (aglomerados);

j. Atribuir a cada protótipo um voto ponderado que represente um artifício para validação de seu erro;

k. Considerar, como a predição da “ensemble”, o rótulo (que está associado ao seu respectivo modelo) que obtiver a maior quantidade de votos.

2. PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES, de acordo com a reivindicação 1, caracterizado por utilizar em sua etapa “c” árvores com aumento de gradiente (XGBoost) e a técnica “Random Forests” como algoritmos de aprendizagem.

3. PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES, de acordo com a reivindicação 1, caracterizado por realizar em sua etapa “g” a maximização da distância euclidiana entre “clusters” em que se encontram os modelos dentro do espaço de modelos H′, inclusive promovendo a maximização da distância euclidiana entre “clusters” por meio da escolha e definição de parâmetros concernentes ao processo de clusterização.

4. PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES, de acordo com a reivindicação 1, caracterizado por utilizar em sua etapa “g” como critério de clusterização de modelos os fatores de explicação denotados pelos valores SHAP (“SHapley Additive exPlanations”) ou outra métrica baseada nos valores SHAP dos modelos.

5. PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS

EXPLICAÇÕES, de acordo com a reivindicação 1, caracterizado por utilizar em sua etapa “g” valores SHAP (“SHapley Additive exPlanations”) para avaliar a importância das características (“features”) na predição do modelo.

6. PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES, de acordo com a reivindicação 1, caracterizado por utilizar em sua etapa “g” como critério de formação dos aglomerados a distância entre seus vetores de explicação e, para maximizar a coesão e separação dos “clusters”, utilizar um critério com base na distância envolvendo a preferência de modelo.

7. USO DO PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES definido na reivindicação 1, caracterizado por utilizar escalas de avaliação de dor como rótulos e base para definição de “features” na criação de modelos de Aprendizado de Máquina para predição acerca da evolução do alívio da dor.

DESENHOS

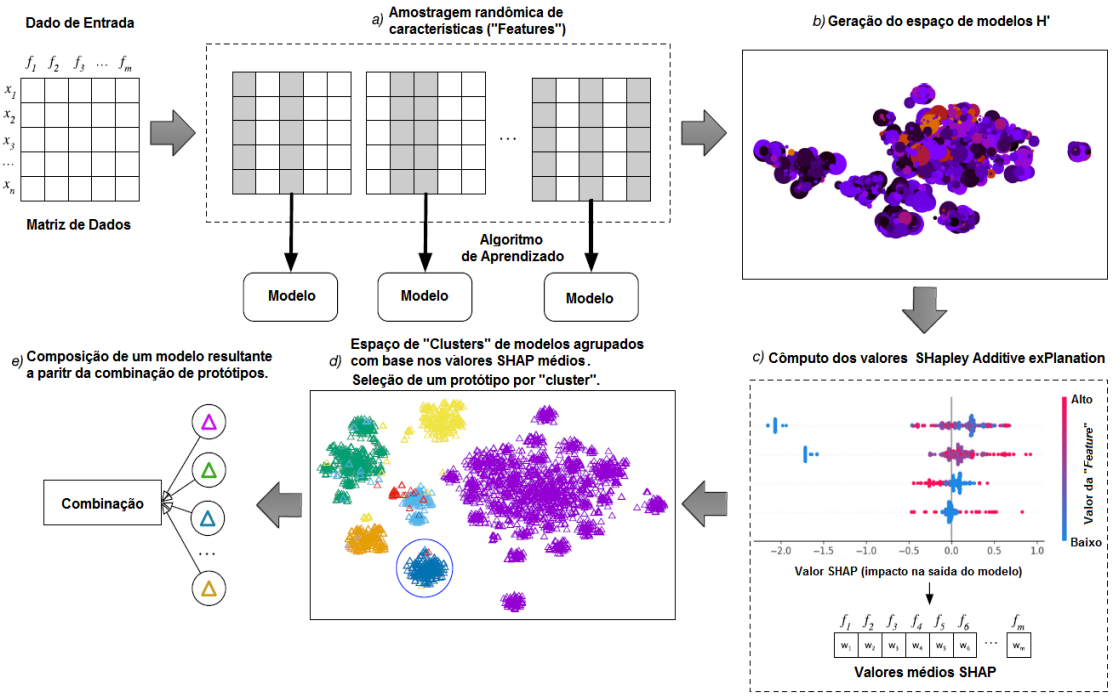

FIGURA 1

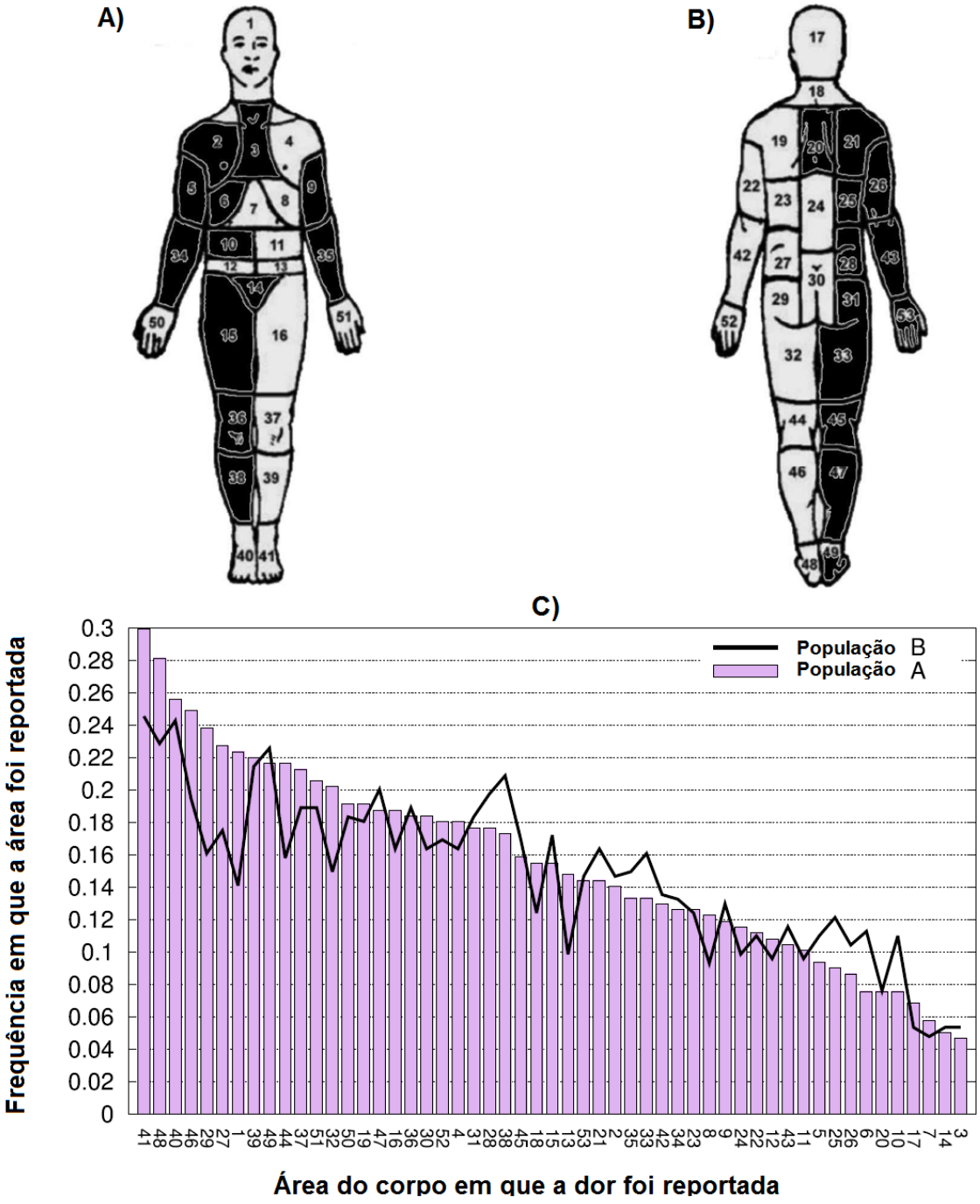
A)
B)
C)
Frequência em que a área foi reportada
0.3
0.28
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
População B
População A
Área do corpo em que a dor foi reportada

FIGURA 2

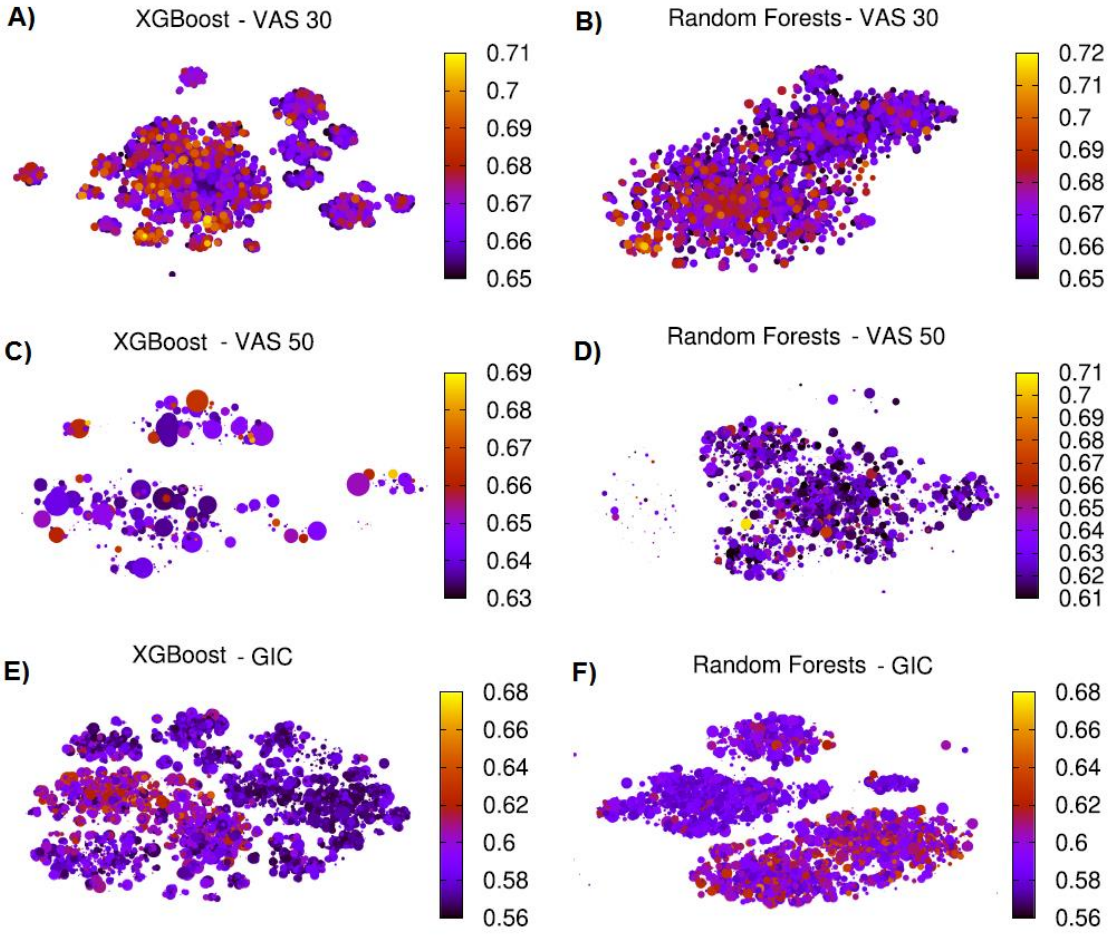
A)
XGBoost - VAS 30
0.71
0.7
0.69
0.68
0.67
0.66
0.65
B)
Random Forests - VAS 30
0.72
0.71
0.7
0.69
0.68
0.67
0.66
0.65
C)
XGBoost - VAS 50
0.69
0.68
0.67
0.66
0.65
0.64
0.63
D)
Random Forests - VAS 50
0.71
0.7
0.69
0.68
0.67
0.66
0.65
0.64
0.63
0.62
0.61
E)
XGBoost - GIC
0.68
0.66
0.64
0.62
0.6
0.58
0.56
F)
Random Forests - GIC
0.68
0.66
0.64
0.62
0.6
0.58
0.56

FIGURA 3

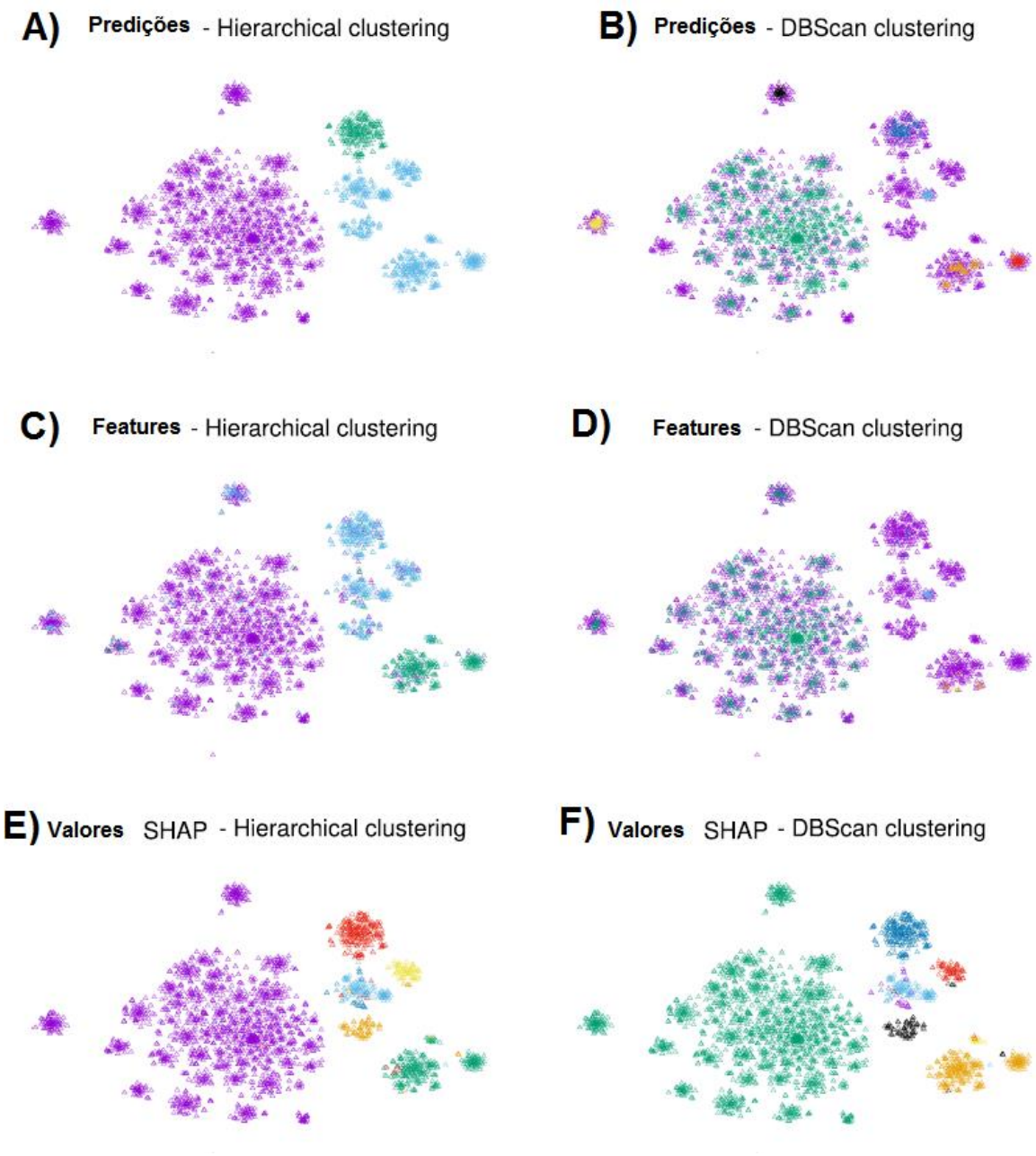
A) Predições - Hierarchical clustering
B) Predições - DBScan clustering
C) Features - Hierarchical clustering
D) Features - DBScan clustering
E) Valores SHAP - Hierarchical clustering
F) Valores SHAP - DBScan clustering

FIGURA 4

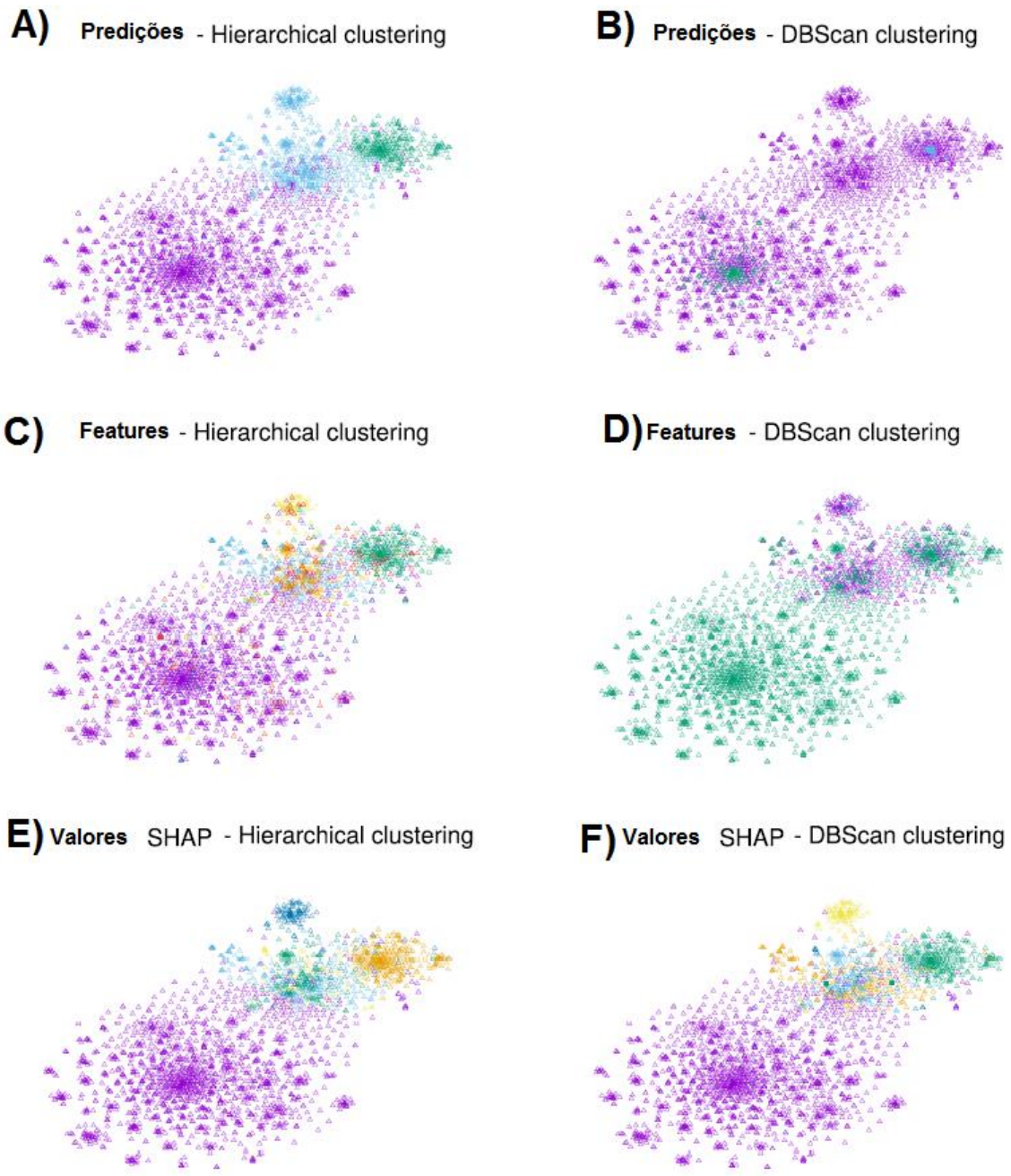
A) Predições - Hierarchical clustering
B) Predições - DBScan clustering
C) Features - Hierarchical clustering
D) Features - DBScan clustering
E) Valores SHAP - Hierarchical clustering
F) Valores SHAP - DBScan clustering

FIGURA 5

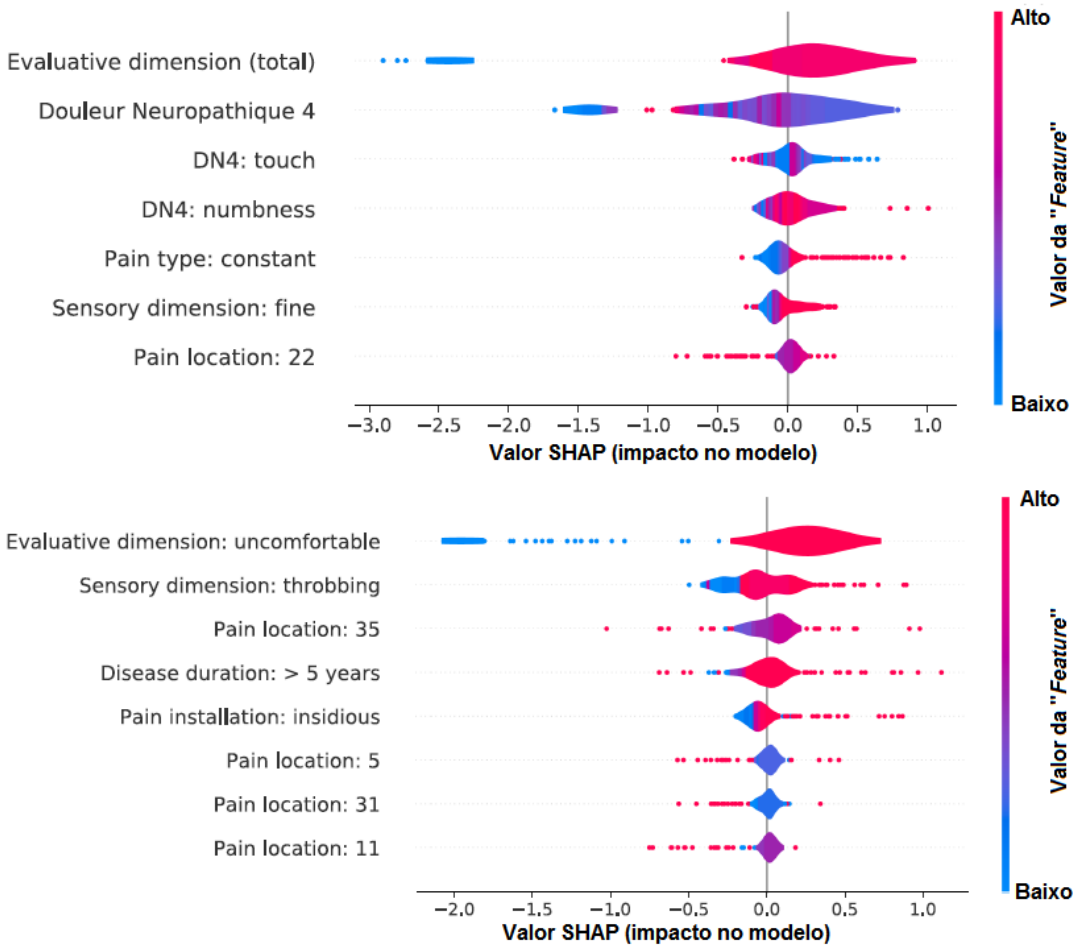
Evaluative dimension (total)
Douleur Neuropathique 4
DN4: touch
DN4: numbness
Pain type: constant
Sensory dimension: fine
Pain location: 22
−3.0 −2.5 −2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0
Valor SHAP (impacto no modelo)
Alto
Valor da "Feature"
Baixo
Evaluative dimension: uncomfortable
Sensory dimension: throbbing
Pain location: 35
Disease duration: > 5 years
Pain installation: insidious
Pain location: 5
Pain location: 31
Pain location: 11
−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0
Valor SHAP (impacto no modelo)
Alto
Valor da "Feature"
Baixo

FIGURA 6

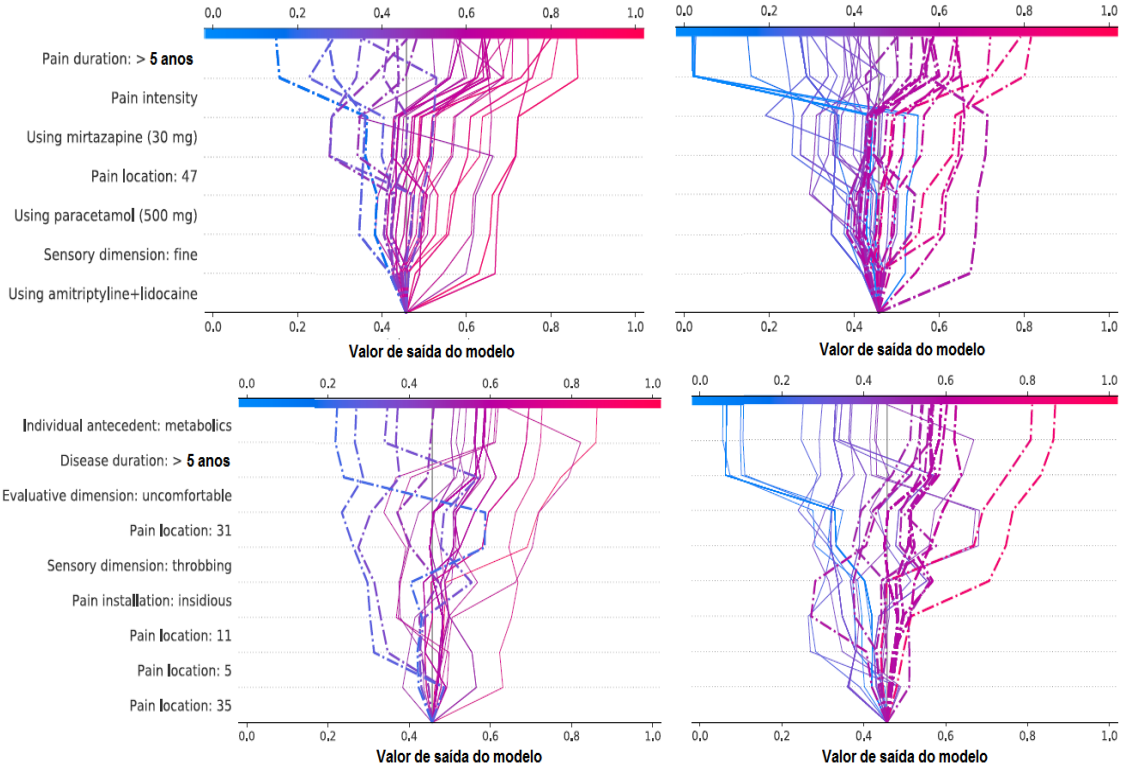
Pain duration: > 5 anos
Pain intensity
Using mirtazapine (30 mg)
Pain location: 47
Using paracetamol (500 mg)
Sensory dimension: fine
Using amitriptyline+lidocaine
Individual antecedent: metabolics
Disease duration: > 5 anos
Evaluative dimension: uncomfortable
Pain location: 31
Sensory dimension: throbbing
Pain installation: insidious
Pain location: 11
Pain location: 5
Pain location: 35
0.0
0.2
0.4
0.6
0.8
1.0
Valor de saída do modelo

FIGURA 7

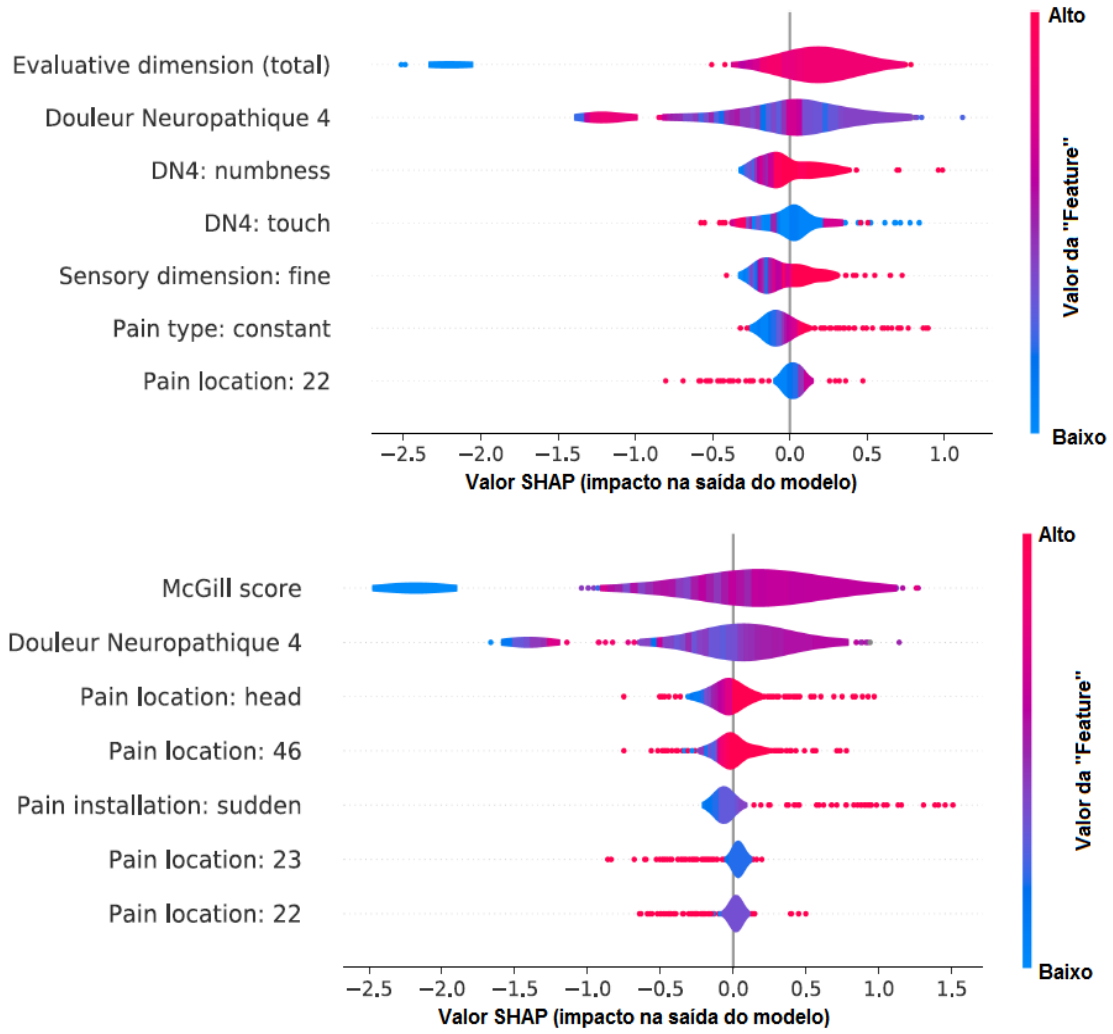
Evaluative dimension (total)
Douleur Neuropathique 4
DN4: numbness
DN4: touch
Sensory dimension: fine
Pain type: constant
Pain location: 22
−2.5 −2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0
Valor SHAP (impacto na saída do modelo)
Alto
Valor da "Feature"
Baixo
McGill score
Douleur Neuropathique 4
Pain location: head
Pain location: 46
Pain installation: sudden
Pain location: 23
Pain location: 22
−2.5 −2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5
Valor SHAP (impacto na saída do modelo)
Alto
Valor da "Feature"
Baixo

FIGURA 8

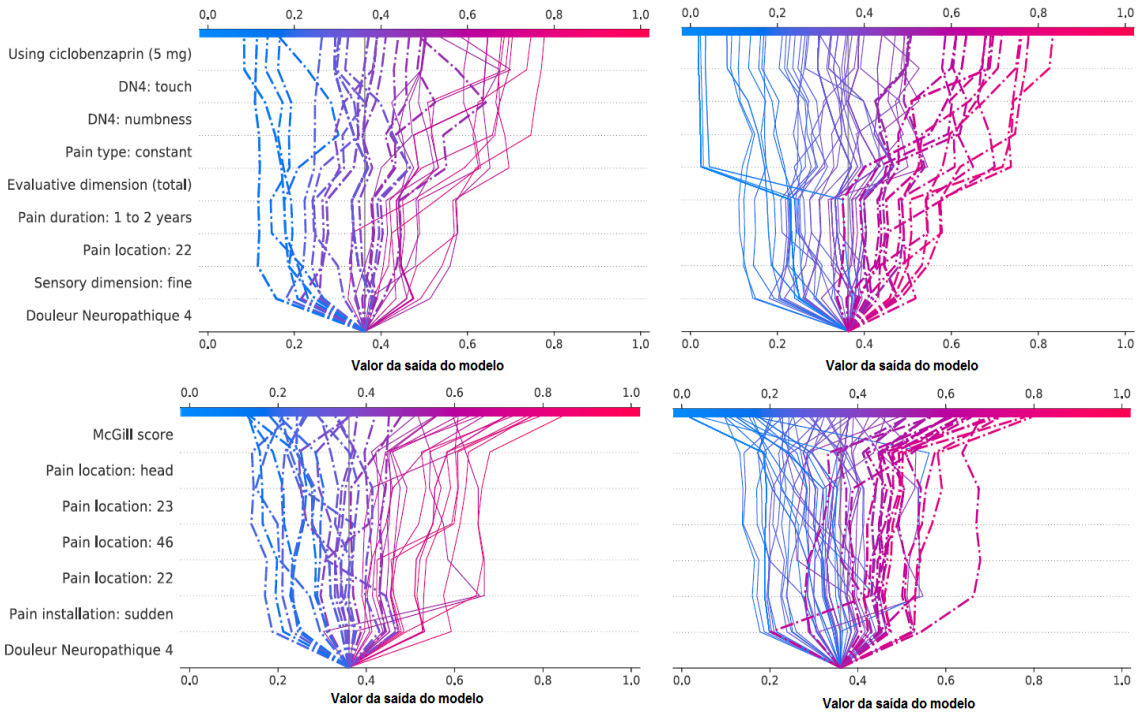
Using ciclobenzaprin (5 mg)
DN4: touch
DN4: numbness
Pain type: constant
Evaluative dimension (total)
Pain duration: 1 to 2 years
Pain location: 22
Sensory dimension: fine
Douleur Neuropathique 4
0.0
0.2
0.4
0.6
0.8
1.0
Valor da saída do modelo
McGill score
Pain location: head
Pain location: 23
Pain location: 46
Pain location: 22
Pain installation: sudden
Douleur Neuropathique 4
Valor da saída do modelo

FIGURA 9

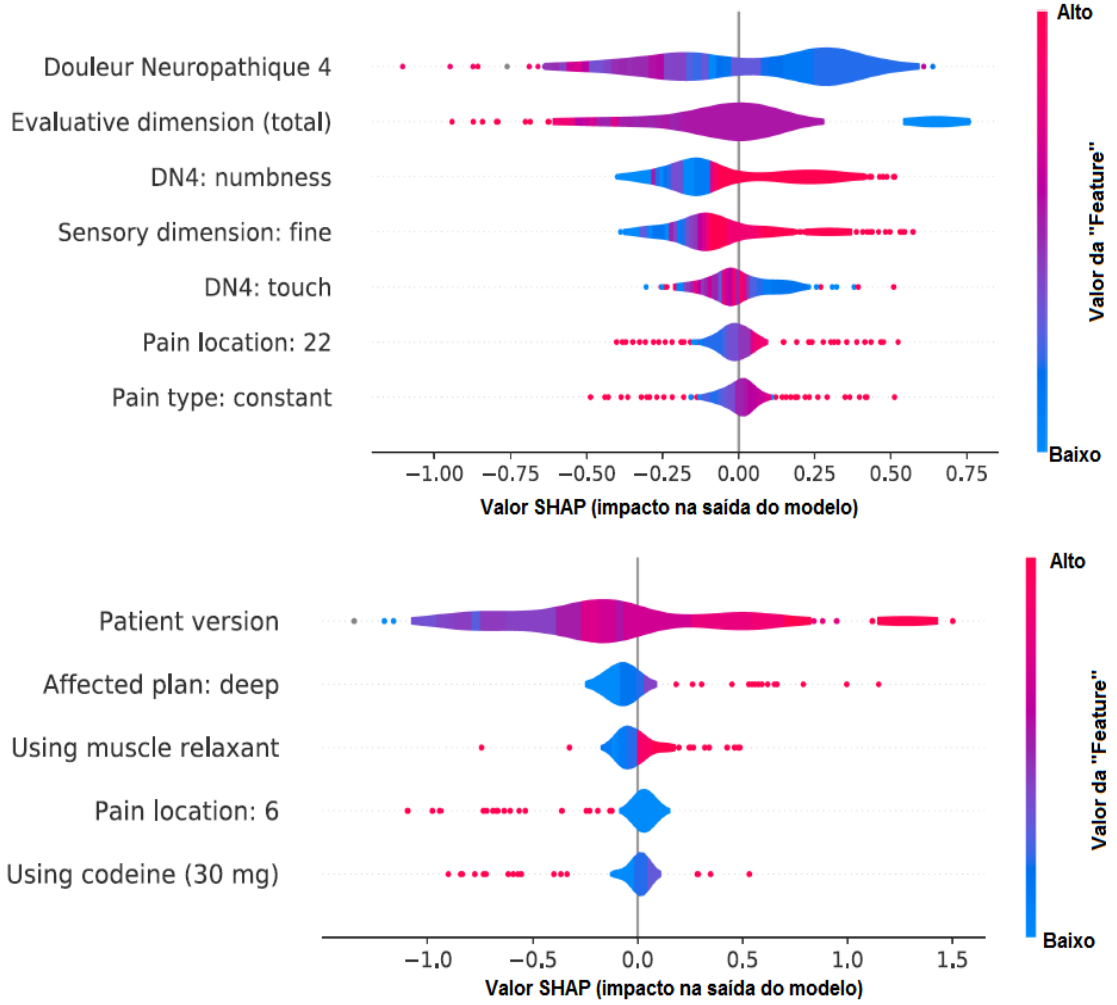
Douleur Neuropathique 4
Evaluative dimension (total)
DN4: numbness
Sensory dimension: fine
DN4: touch
Pain location: 22
Pain type: constant
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75
Valor SHAP (impacto na saída do modelo)
Alto
Valor da "Feature"
Baixo
Patient version
Affected plan: deep
Using muscle relaxant
Pain location: 6
Using codeine (30 mg)
−1.0 −0.5 0.0 0.5 1.0 1.5
Valor SHAP (impacto na saída do modelo)
Alto
Valor da "Feature"
Baixo

FIGURA 10

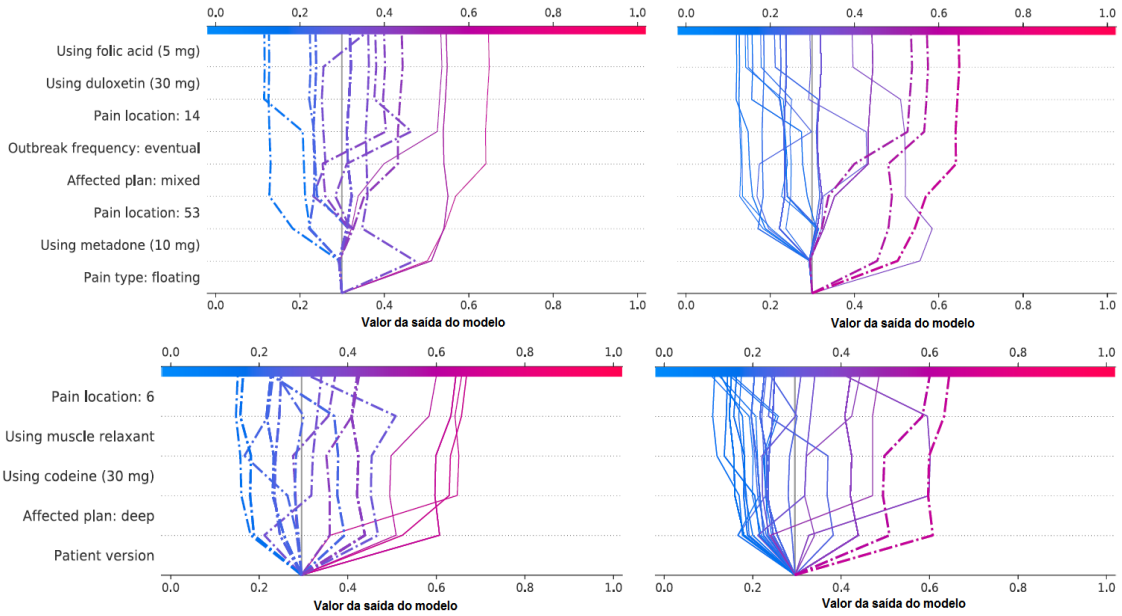

FIGURA 11

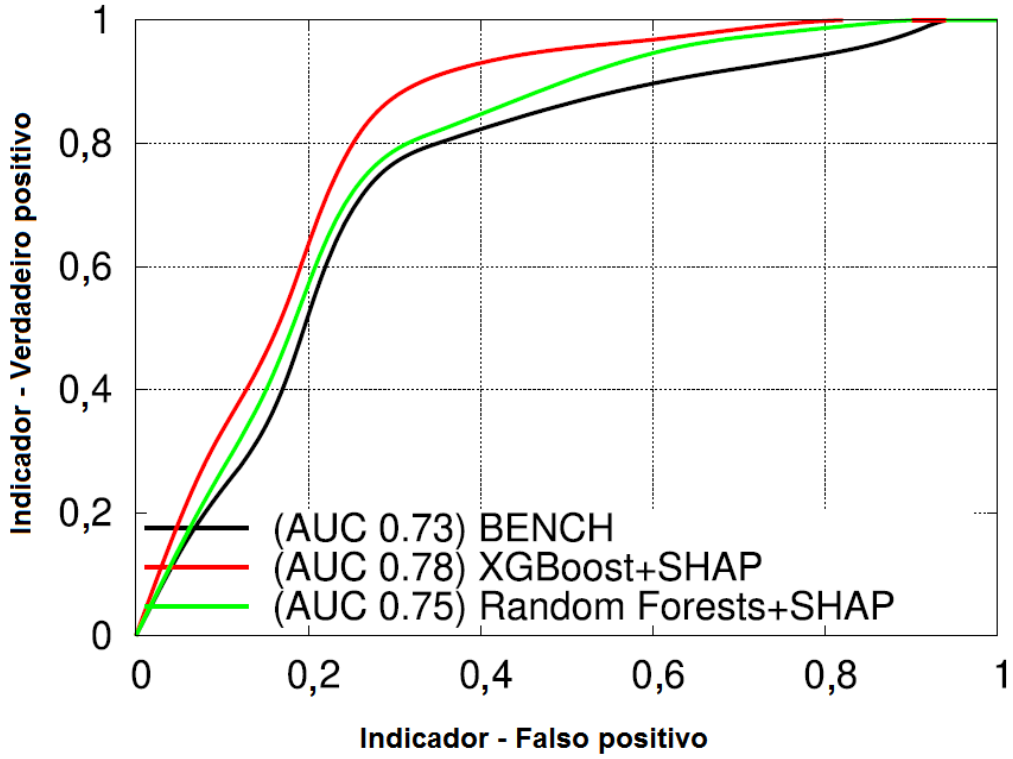

FIGURA 12

RESUMO

“PROCESSO PARA ELABORAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA BASEADO NA DIVERSIFICAÇÃO DAS EXPLICAÇÕES E USO”

Esta tecnologia refere-se a um processo de obtenção de modelos baseados em Aprendizado de Máquina que inclui meios para promover a diversificação das explicações do modelo. A tecnologia utiliza técnicas como a decomposição do espaço de dados em estruturas locais, do tipo “*backbone*”, formando um conjunto do tipo “*backbone features*”; representações de explicações de modelos (explicabilidade) utilizadas como critério de clusterização e seleção de protótipos dentro do espaço de modelos para promover a diversificação de explicações. Utiliza também combinação (“ensemble”) de modelos, dentre outras técnicas. As vantagens propiciadas pela tecnologia são principalmente: 1) agilidade na modelagem devido à criação de características relevantes ao problema, economizando tempo e recursos computacionais, favorecendo os ajustes de hiperparâmetros; 2) versatilidade, já que a metodologia pode ser utilizada em problemas de qualquer segmento e formato de banco de dados; 3) propicia o entendimento dos modelos devido à utilização de técnicas de explicabilidade; 4) viabiliza a diversificação das explicações do modelo. A tecnologia aplica-se no contexto de solução de problemas em que se recorre a modelos baseados em Aprendizado de Máquina, para aperfeiçoar a elaboração e o desempenho de tais modelos.

ACM Transactions on Computing for Healthcare

RESEARCH-ARTICLE

Predicting the Evolution of Pain Relief: Ensemble Learning by Diversifying Model Explanations

Authors: Anderson Bessa Da Costa, Larissa Moreira, Daniel Ciampi De Andrade, Adriano Veloso, Nivio Ziviani Authors Info & Claims

ACM Transactions on Computing for Healthcare, Volume 2, Issue 4 • October 2021 • Article No.: 36, pp 1–28 • https://doi.org/10.1145/3466781

Online: 14 September 2021 Publication History

Abstract

Modeling from data usually has two distinct facets: building sound explanatory models or c powerful predictive models for a system or phenomenon. Most of recent literature does not exploit the relationship between explanation and prediction while learning models from data. Recent algorithms are not taking advantage of the fact that many phenomena are actually defined by diverse sub-populations and local structures, and thus there are many possible predictive models providing contrasting interpretations or competing explanations for the same phenomenon. In this article, we propose to explore a complementary link between explanation

evaluate our methodology to model the evolution of pain relief in patients suffering from chronic

Random Forests and XGBoost. Chronic pain can be primary or secondary to diseases. Its symptomatology can be classified as nociceptive, nociplastic, or neuropathic, and is generally associated with many different causal structures, challenging typical modeling methodologies. Our data includes 631 patients receiving pain treatment. We considered 338 features providing information about pain sensation, socioeconomic status, and prescribed treatments. Our goal is to predict, using data from the first consultation only, if the patient will be successful in treatment for chronic pain relief. As a result of this work, we were able to build ensembles that are able to consistently improve performance by up to 33% when compared to models trained using all the available features. We also obtained relevant gains in interpretability, with resulting ensembles using only 15% of the total number of features. We show we can effectively generate ensembles from competing explanations, promoting diversity in ensemble learning and leading to significant gains in accuracy by enforcing a stable scenario in which models that are dissimilar in terms of their predictions are also dissimilar in terms of their explanation factors.

References

1. M. Abad-Grau, J. Ierache, C. Cervino, and P. Sebastiani. 2008. Evolution and challenges in the design of computational systems for triage assistance. J. Biomed. Inform. 41, 3 (2008), 432–441.

2. R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. 2005. Automatic subspace clustering of high dimensional data. Data Mining Knowl. Discov. 11, 1 (2005), 5–33.

3. J. M. Benitez, J. L. Castro, and I. Requena. 1997. Are artificial neural networks black boxes?IEEE Trans. Neural Netw. 8, 5 (Sept. 1997), 1156–1164. DOI:https://doi.org/10.1109/72.623216

Show All References

Index Terms

Applied computing

Computing methodologies

Life and medical sciences

Health informatics

Machine learning

Machine learning algorithms

Ensemble methods

Comments

DL Comment Policy

Comments should be relevant to the contents of this article, (sign in required).

0 Comments

Tweet Share

Sort by Newest

Nothing in this discussion yet.

View Issue's Table of Contents