



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

Número do Processo: BR 10 2021 015411 0

### 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

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 Professor José Vieira de Mendonça, 770 sala 208 - Engenho Nogueira

**Cidade:** Belo Horizonte

**Estado:** MG

**CEP:** 31310-260

**País:** BRASIL

**Telefone:**

**Fax:**

**Email:**

**Natureza Patente:** 10 - Patente de Invenção (PI)

**Título da Invenção ou Modelo de Utilidade (54):** PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA E USOS

**Resumo:** A presente tecnologia se refere a um processo centrado no humano para obtenção de modelos baseados em aprendizado de máquina que inclui meios para permitir a um especialista (humano) contribuir para elaboração de modelos, julgando-os. A tecnologia utiliza técnicas como a decomposição do espaço de dados em estruturas locais, do tipo backbone, formando um conjunto do tipo aka backbone features; representações vetoriais de explicações de modelos (explicabilidade); agrupamento (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 contribuição de especialista humano ao modelo, utilizando metáforas de explicação para julgar características do modelo potencialmente problemáticas, ou até mesmo identificar a ausência de características indispensáveis à modelagem. A tecnologia se aplica 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.

## Dados do Inventor (72)

---

### Inventor 1 de 6

**Nome:** ADRIANO ALONSO VELOSO

**CPF:** 03733647688

**Nacionalidade:** Brasileira

**Qualificação Física:** Professor do ensino superior

**Endereço:** Av. Pres. Antônio Carlos, 6627 - Pampulha

**Cidade:** Belo Horizonte

**Estado:** MG

**CEP:** 31270-901

**País:** BRASIL

**Telefone:** (31) 340 93932

**Fax:**

**Email:** patentes@ctit.ufmg.br

### Inventor 2 de 6

**Nome:** PAULO CARAMELLI

**CPF:** 05095003821

**Nacionalidade:** Brasileira

**Qualificação Física:** Professor do ensino superior

**Endereço:** Av. Pres. Antônio Carlos, 6627 - Pampulha

**Cidade:** Belo Horizonte

**Estado:** MG

**CEP:** 31270-901

**País:** BRASIL

**Telefone:** (31) 340 93932

**Fax:**

**Email:** patentes@ctit.ufmg.br

### Inventor 3 de 6

**Nome:** KARINA BRAGA GOMES BORGES

**CPF:** 03132898678

**Nacionalidade:** Brasileira

**Qualificação Física:** Professor do ensino superior

**Endereço:** Av. Pres. Antônio Carlos, 6627 - Pampulha

**Cidade:** Belo Horizonte

**Estado:** MG

**CEP:** 31270-901

**País:** BRASIL

**Telefone:** (31) 340 93932

**Fax:**

**Email:** patentes@ctit.ufmg.br

**Inventor 4 de 6**

**Nome:** DANIELLA CASTRO DE ARAÚJO

**CPF:** 71989536115

**Nacionalidade:** Brasileira

**Qualificação Física:** Pesquisador

**Endereço:** Av. Pres. Antônio Carlos, 6627 - Pampulha

**Cidade:** Belo Horizonte

**Estado:** MG

**CEP:** 31270-901

**País:** BRASIL

**Telefone:** (31) 340 93932

**Fax:**

**Email:** patentes@ctit.ufmg.br

**Inventor 5 de 6**

**Nome:** GIANLUCCA ZUIN

**CPF:** 12358810673

**Nacionalidade:** Brasileira

**Qualificação Física:** Pesquisador

**Endereço:** Av. Pres. Antônio Carlos, 6627 - Pampulha

**Cidade:** Belo Horizonte

**Estado:** MG

**CEP:** 31270-901

**País:** BRASIL

**Telefone:** (31) 340 93932

**Fax:**

**Email:** patentes@ctit.ufmg.br

**Inventor 6 de 6**

**Nome:** TIAGO HENRIQUE COSTA ALVES

**CPF:** 11705343686

**Nacionalidade:** Brasileira

**Qualificação Física:** Pesquisador

**Endereço:** Av. Pres. Antônio Carlos, 6627 - Pampulha

**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 - Comprovante de pagamento GRU - 29409161935218297-mesclado.pdf
Portaria	2 - Portaria 2195-2020 - Prof. Gilberto UFMG.pdf
Procuração	3 - Procuração KUNUMI para UFMG - NI 118.2019 (1).docx (3).pdf
Comprovação de poderes	4 - Comprovação de poderes Kunumi.pdf
Relatório Descritivo	5 - Relatório descritivo.pdf
Reivindicação	6 - Reivindicações.pdf
Resumo	7 - Resumo.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 veracidade

---

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

**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 35218.297170 7 86480000007000

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				
Noosso-Número	Nr. Documento	Data de Vencimento	Valor do Documento	(=) Valor Pago
29409161935218297	29409161935218297	11/06/2021	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			Autenticação Mecânica	
2234-9 / 333028-1				

**BANCO DO BRASIL** | 001-9 | 00190.00009 02940.916196 35218.297170 7 86480000007000

Local de Pagamento					Data de Vencimento	
<b>PAGÁVEL EM QUALQUER BANCO ATÉ O VENCIMENTO</b>					11/06/2021	
Nome do Beneficiário/CPF/CNPJ					Agência/Código do Beneficiário	
INSTITUTO NACIONAL DA PROPRIEDADE INDUST CPF/CNPJ: 42.521.088/0001-37					2234-9 / 333028-1	
Data do Documento	Nr. Documento	Espécie DOC	Aceite	Data do Processamento	Nosso-Número	
13/05/2021	29409161935218297	DS	N	13/05/2021	29409161935218297	
Uso do Banco	Carteira	Espécie	Quantidade	xValor	(=) Valor do Documento	
29409161935218297	17	R\$			70,00	
Informações de Responsabilidade do Beneficiário					(-) Desconto/Abatimento	
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					(+ Juros/Multa	
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					(-) Valor Cobrado	

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



\_\_\_ SIAFI2021-DOCUMENTO-CONSULTA-CONGRU (CONSULTA GUIA DE RECOLHIMENTO DA UNIAO  
19/05/21 13:04 USUARIO : LUDMILA  
DATA EMISSAO : 19Mai21 TIPO : 1 - PAGAMENTO NUMERO : 2021GR800477  
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: MAI21 VENCIMENTO: 19Mai21  
DOC. ORIGEM: 153254 / 15229 / 2021NP000954 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 : 00029409161935218297  
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 19Mai2021 12:44  
PF1=AJUDA PF3=SAI PF2=DADOS ORC/FIN PF4=ESPELHO PF12=RETORNA



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](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**.

## PROCURAÇÃO

Por este instrumento particular de Procuração, a Kunumi Serviços Em Tecnologia da Informação S/A, com sede na Rua Professor José Vieira de Mendonça, 770 sala 208 - Engenho Nogueira, Belo Horizonte, Minas Gerais, inscrita no CNPJ sob o nº 24.477.718/0001-31, neste ato representada por seus diretores Maria Gabriella Grabowsky Seiler- diretora Acadêmica e Pesquisa e Maurício Campos Zuardi- Diretor de Tecnologia, 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 face ao pedido de patente intitulado “**PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA E USOS**”, a ser depositado junto ao INPI, e assim 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, 04 de agosto de 2021.

---

Maria Gabriella Grabowsky Seiler

---

Maurício Campos Zuardi



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)

31300114333

Código da Natureza Jurídica

2054

Nº de Matrícula do Agente Auxiliar do Comércio

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

Processo em Ordem À decisão

\_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_  
Data

NÃO

\_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_  
Data

\_\_\_\_\_  
Responsável

NÃO

\_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_  
Data

\_\_\_\_\_  
Responsável

\_\_\_\_\_  
Responsável

DECISÃO SINGULAR

Processo em exigência. (Vide despacho em folha anexa)

Processo deferido. Publique-se e archive-se.

Processo indeferido. Publique-se.

2ª Exigência

3ª Exigência

4ª Exigência

5ª Exigência

\_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_  
Data

\_\_\_\_\_  
Responsável

DECISÃO COLEGIADA

Processo em exigência. (Vide despacho em folha anexa)

Processo deferido. Publique-se e archive-se.

Processo indeferido. Publique-se.

2ª Exigência

3ª Exigência

4ª Exigência

5ª Exigência

\_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_  
Data

\_\_\_\_\_  
Vogal

\_\_\_\_\_  
Vogal

\_\_\_\_\_  
Vogal

Presidente da \_\_\_\_\_ Turma

OBSERVAÇÕES



Junta Comercial do Estado de Minas Gerais

Certifico registro sob o nº 7855420 em 29/05/2020 da Empresa KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A, Nire 31300114333 e protocolo 203052404 - 28/05/2020. Autenticação: BACAC1437794C5313FEBC57678BFC188609C8D16. Marinely de Paula Bomfim - Secretária-Geral. Para validar este documento, acesse <http://www.jucemg.mg.gov.br> e informe nº do protocolo 20/305.240-4 e o código de segurança kp2G

Esta cópia foi autenticada digitalmente e assinada em 29/05/2020 por Marinely de Paula Bomfim Secretária-Geral.

MARINELY DE PAULA BOMFIM  
SECRETÁRIA GERAL



# JUNTA COMERCIAL DO ESTADO DE MINAS GERAIS

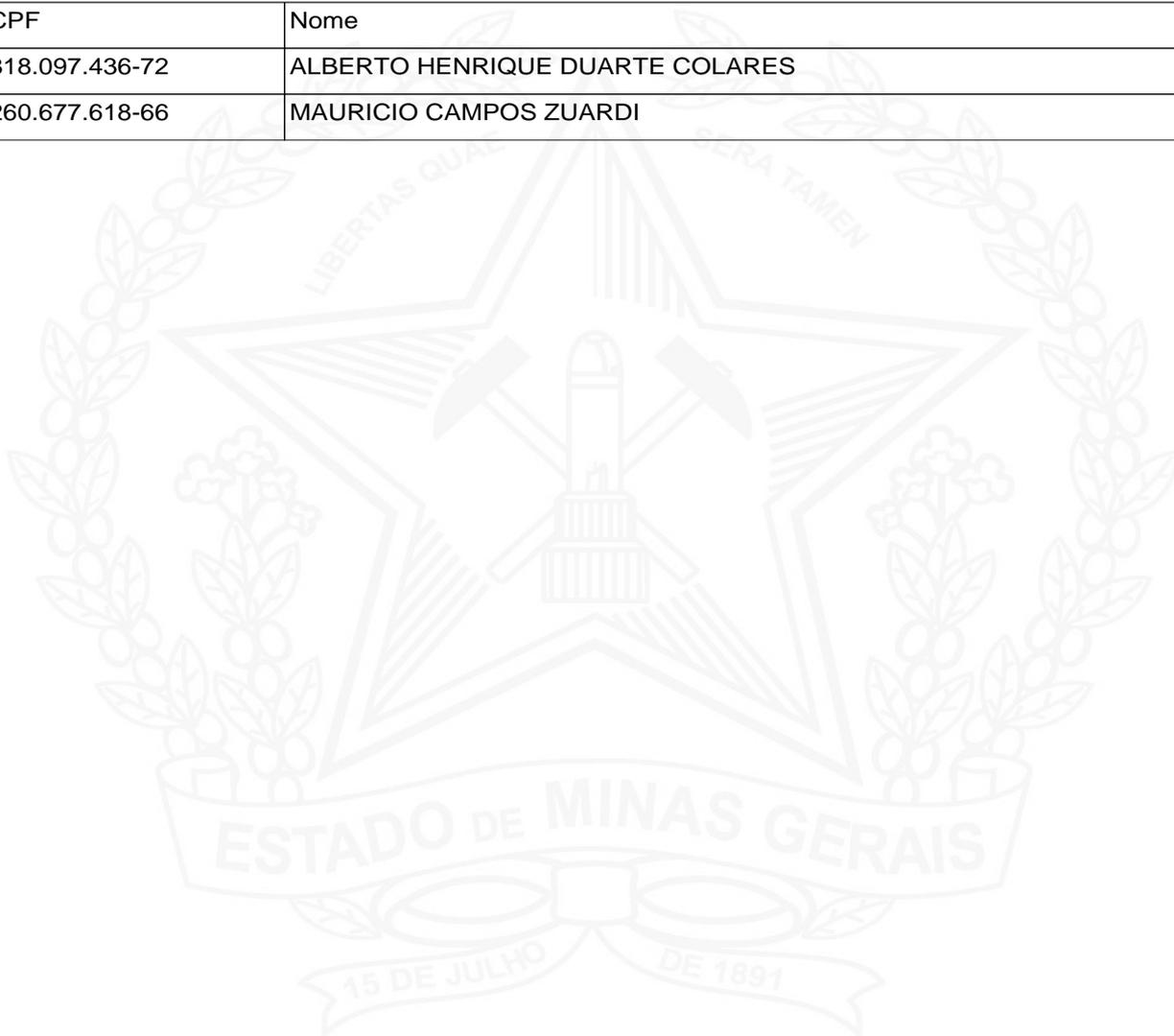
Registro Digital

## 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

Junta Comercial do Estado de Minas Gerais



**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 PRESENCAS:** 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 CPF/MF 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**

Ana Paula Machado Pessoa

*Assinado por meio de certificado digital*

**Secretário**

Ramon Dias de Azevedo

*Assinado por meio de certificado digital*





# JUNTA COMERCIAL DO ESTADO DE MINAS GERAIS

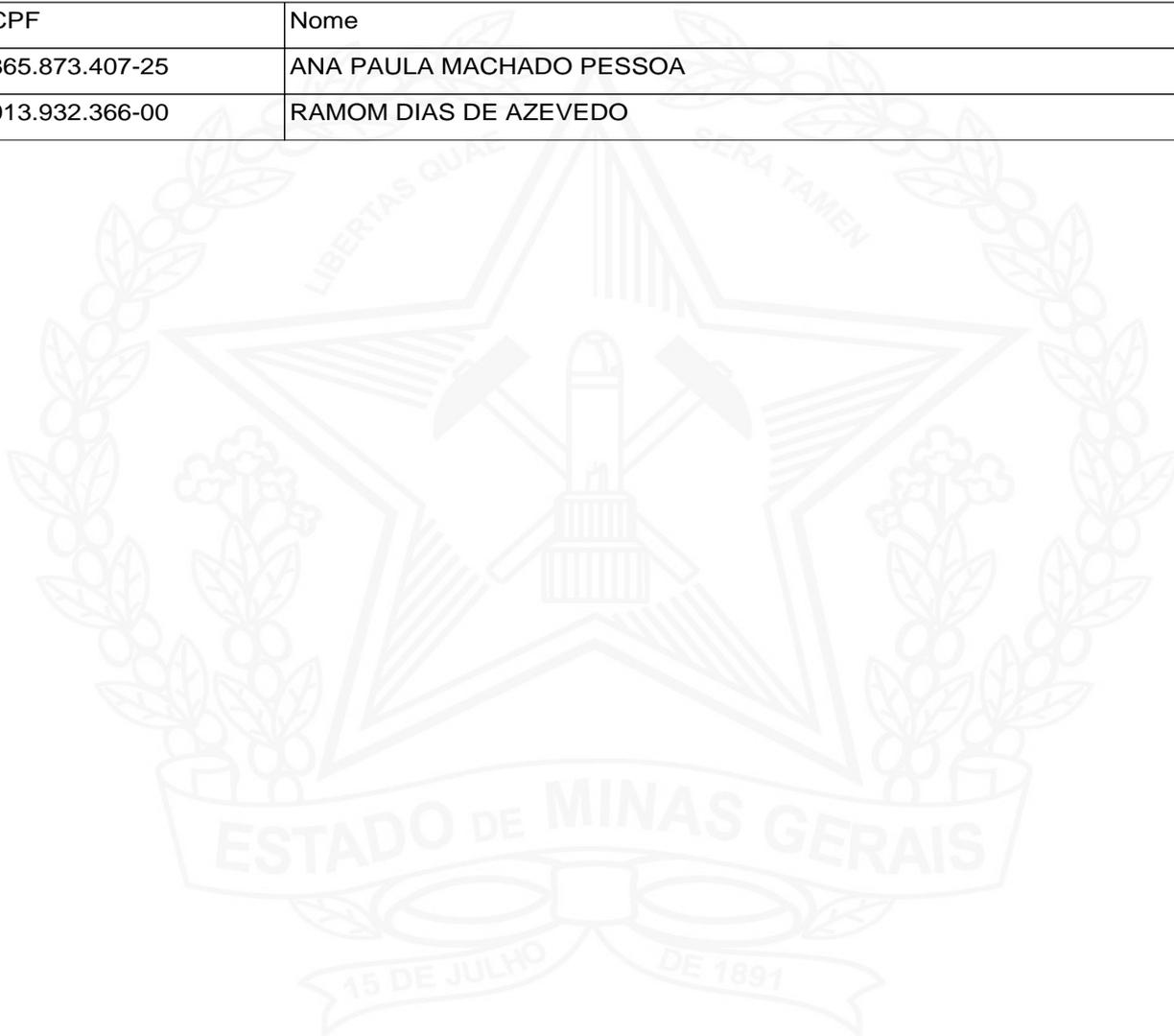
Registro Digital

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

Junta Comercial do Estado de Minas Gerais



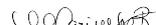
Página 1 de 1



Junta Comercial do Estado de Minas Gerais

Certifico registro sob o nº 7855420 em 29/05/2020 da Empresa KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A, Nire 31300114333 e protocolo 203052404 - 28/05/2020. Autenticação: BACAC1437794C5313FEBC57678BFC188609C8D16. Marinely de Paula Bomfim - Secretária-Geral. Para validar este documento, acesse <http://www.jucemg.mg.gov.br> e informe nº do protocolo 20/305.240-4 e o código de segurança kp2G. Esta cópia foi autenticada digitalmente e assinada em 29/05/2020 por Marinely de Paula Bomfim Secretária-Geral.

Petição 870210071010, de 04/08/2021, pag. 18/64

  
MARINELY DE PAULA BOMFIM  
SECRETARIA GERAL

pág. 6/17

## 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**



Junta Comercial do Estado de Minas Gerais

Certifico registro sob o nº 7855420 em 29/05/2020 da Empresa KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A, Nire 31300114333 e protocolo 203052404 - 28/05/2020. Autenticação: BACAC1437794C5313FEB57678BFC188609C8D16. Marinely de Paula Bomfim - Secretária-Geral. Para validar este documento, acesse <http://www.jucemg.mg.gov.br> e informe nº do protocolo 20/305.240-4 e o código de segurança kp2G



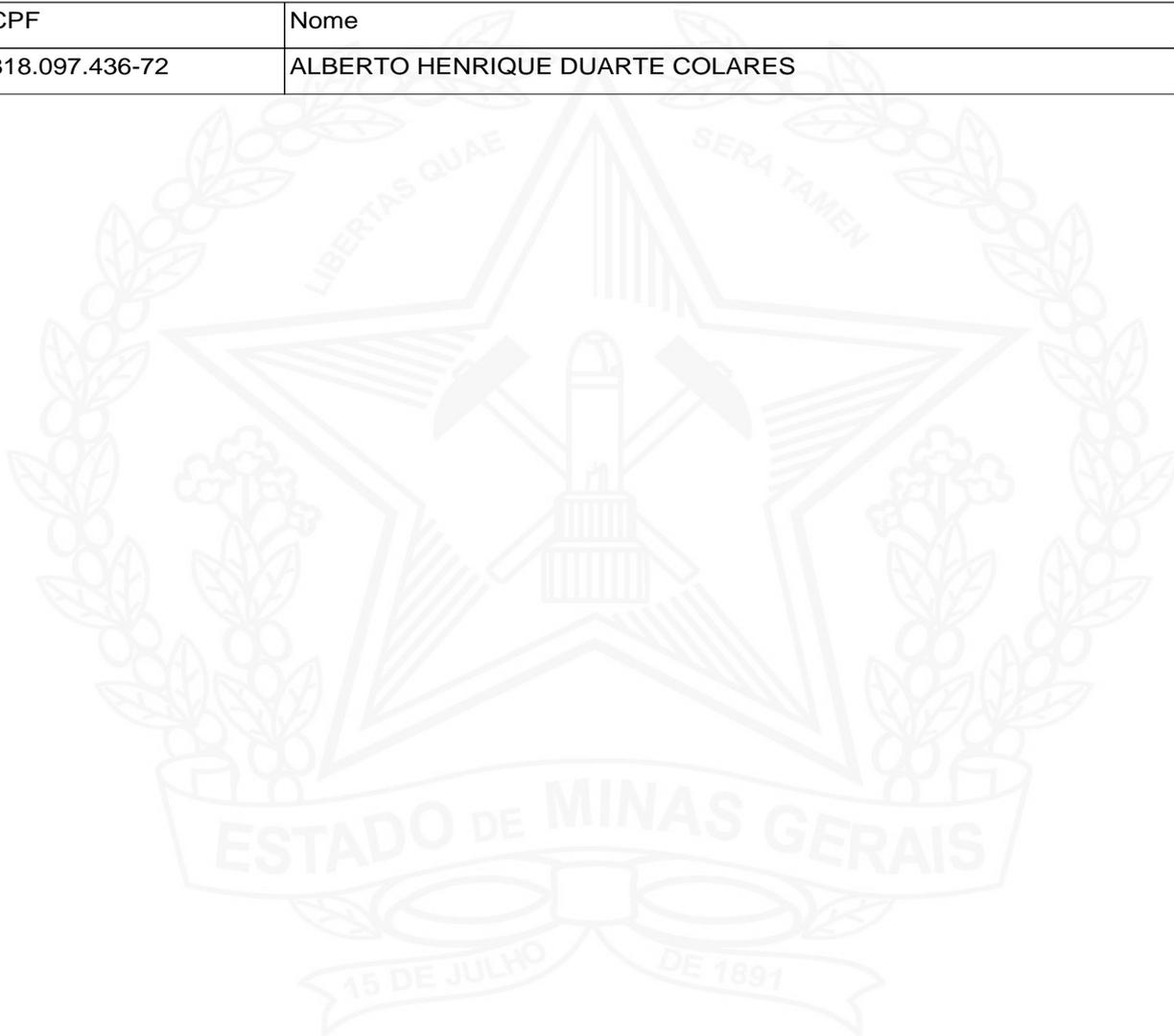
# JUNTA COMERCIAL DO ESTADO DE MINAS GERAIS

Registro Digital

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



Junta Comercial do Estado de Minas Gerais

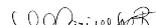
Página 1 de 1



Junta Comercial do Estado de Minas Gerais

Certifico registro sob o nº 7855420 em 29/05/2020 da Empresa KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A, Nire 31300114333 e protocolo 203052404 - 28/05/2020. Autenticação: BACAC1437794C5313FEB57678BFC188609C8D16. Marinely de Paula Bomfim - Secretária-Geral. Para validar este documento, acesse <http://www.jucemg.mg.gov.br> e informe nº do protocolo 20/305.240-4 e o código de segurança kp2G

Esta cópia foi autenticada digitalmente e assinada em 29/05/2020 por Marinely de Paula Bomfim Secretária-Geral.

  
MARINELY DE PAULA BOMFIM  
SECRETÁRIA GERAL

pág. 8/17

## 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 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.

---

**MARIA GABRIELLA GRABOWSKY SEILER**



Junta Comercial do Estado de Minas Gerais

Certifico registro sob o nº 7855420 em 29/05/2020 da Empresa KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A, Nire 31300114333 e protocolo 203052404 - 28/05/2020. Autenticação: BACAC1437794C5313FEB57678BFC188609C8D16. Marinely de Paula Bomfim - Secretária-Geral. Para validar este documento, acesse <http://www.jucemg.mg.gov.br> e informe nº do protocolo 20/305.240-4 e o código de segurança kp2G



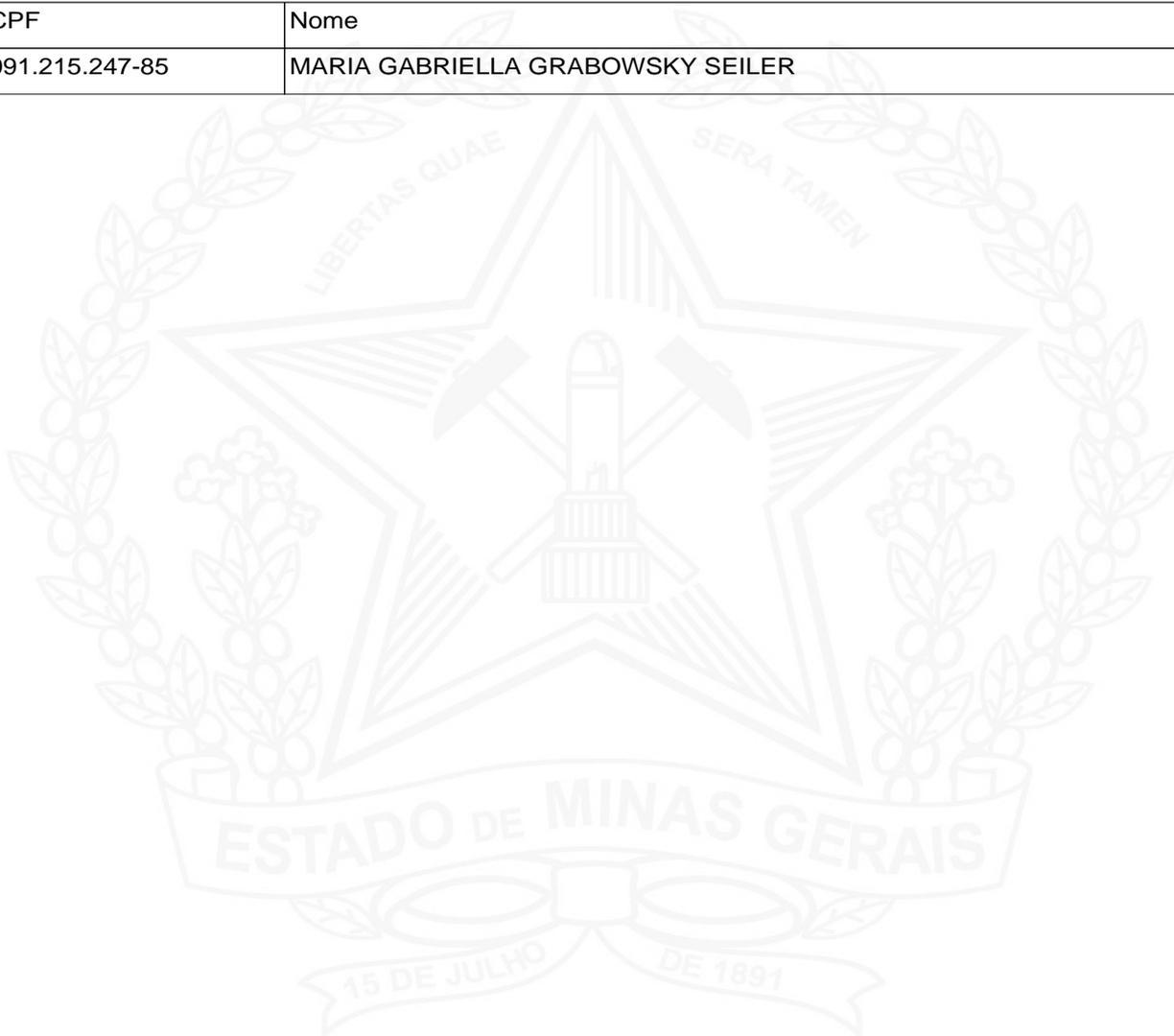
# JUNTA COMERCIAL DO ESTADO DE MINAS GERAIS

Registro Digital

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



Junta Comercial do Estado de Minas Gerais



## 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**



Junta Comercial do Estado de Minas Gerais

Certifico registro sob o nº 7855420 em 29/05/2020 da Empresa KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A, Nire 31300114333 e protocolo 203052404 - 28/05/2020. Autenticação: BACAC1437794C5313FEB57678BFC188609C8D16. Marinely de Paula Bomfim - Secretária-Geral. Para validar este documento, acesse <http://www.jucemg.mg.gov.br> e informe nº do protocolo 20/305.240-4 e o código de segurança kp2G



# JUNTA COMERCIAL DO ESTADO DE MINAS GERAIS

Registro Digital

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



Junta Comercial do Estado de Minas Gerais



## 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





# JUNTA COMERCIAL DO ESTADO DE MINAS GERAIS

Registro Digital

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



Junta Comercial do Estado de Minas Gerais

Página 1 de 1



Junta Comercial do Estado de Minas Gerais

Certifico registro sob o nº 7855420 em 29/05/2020 da Empresa KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A, Nire 31300114333 e protocolo 203052404 - 28/05/2020. Autenticação: BACAC1437794C5313FEB57678BFC188609C8D16. Marinely de Paula Bomfim - Secretária-Geral. Para validar este documento, acesse <http://www.jucemg.mg.gov.br> e informe nº do protocolo 20/305.240-4 e o código de segurança kp2G. Esta cópia foi autenticada digitalmente e assinada em 29/05/2020 por Marinely de Paula Bomfim Secretária-Geral.

Petição 870210071010, de 04/08/2021, pag. 26/64

MARINELY DE PAULA BOMFIM  
SECRETÁRIA-GERAL

pág. 14/17



## 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 eletronicamente 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 autenticidade desse documento pode ser conferida no [portal de serviços da jucemg](https://portalservicos.jucemg.mg.gov.br) 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 eletronicamente por Zulene figueiredo, Servidor(a) Público(a), em 29/05/2020, às 12:26 conforme horário oficial de Brasília.



A autenticidade desse documento pode ser conferida no [portal de serviços da jucemg](http://portal.de.servicos.da.jucemg) informando o número do protocolo 20/305.240-4.

Página 2 de 2





# JUNTA COMERCIAL DO ESTADO DE MINAS GERAIS

Registro Digital

O ato foi deferido e assinado digitalmente por :

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

Junta Comercial do Estado de Minas Gerais



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



Junta Comercial do Estado de Minas Gerais

Certifico registro sob o nº 7855420 em 29/05/2020 da Empresa KUNUMI SERVICOS EM TECNOLOGIA DA INFORMACAO S/A, Nire 31300114333 e protocolo 203052404 - 28/05/2020. Autenticação: BACAC1437794C5313FEB57678BFC188609C8D16. Marinely de Paula Bomfim - Secretária-Geral. Para validar este documento, acesse <http://www.jucemg.mg.gov.br> e informe nº do protocolo 20/305.240-4 e o código de segurança kp2G

Esta cópia foi autenticada digitalmente e assinada em 29/05/2020 por Marinely de Paula Bomfim Secretária-Geral.

MARINELY DE PAULA BOMFIM  
SECRETARIA GERAL

pág. 17/17

## **“PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA E USOS”**

[01] A presente tecnologia se refere a um processo centrado no humano para obtenção de modelos baseados em aprendizado de máquina que inclui meios para permitir a um especialista (humano) contribuir para elaboração de modelos, julgando-os. A tecnologia utiliza técnicas como a decomposição do espaço de dados em estruturas locais, do tipo “*backbone*”, formando um conjunto do tipo “*aka backbone features*”; representações vetoriais de explicações de modelos (explicabilidade); agrupamento (“*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 contribuição de especialista humano ao modelo, utilizando metáforas de explicação para julgar características do modelo potencialmente problemáticas, ou até mesmo identificar a ausência de características indispensáveis à modelagem. A tecnologia se aplica 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] Em uma modelagem de Aprendizado de Máquina, dado um determinado problema, os cientistas de dados precisam utilizar um conjunto de características relacionadas de alguma forma ao problema de interesse. Essas características são combinadas de maneira não trivial por sofisticados algoritmos de aprendizado de máquina, e o

cientista de dados atua no processo ajustando os hiperparâmetros do algoritmo manualmente. Assim, são necessárias grandes quantidades de recursos computacionais, bem como tempo, para se obter modelos adequados para o problema. (X. He, K. Zhao, X. Chu, "Au toML: A Survey of the State-of-the-Art ", in arXiv:1908.00709, 2019.)

[03] Além da quantidade de tempo e de recursos necessários para o desenvolvimento de modelos de aprendizado de máquina, outra dificuldade enfrentada pelos cientistas de dados é a interpretação dos resultados. Entender a relevância de cada uma das características utilizadas no modelo para o resultado final é uma tarefa complexa e muitas vezes intratável, a depender da quantidade de características envolvidas na modelagem. Dessa forma, a metodologia pretende: (1) reduzir a necessidade de execução de múltiplos testes para ajustes de hiperparâmetros de modelos de aprendizado de máquina, diminuindo consideravelmente o tempo gasto no desenvolvimento; (2) produzir modelos auto-explicáveis, já que o entendimento das características mais importantes para a tomada de decisão dos modelos é crucial para sua aplicação em diversos segmentos, como saúde, bioquímica, indústria, educação, etc. (3) produzir modelos mais coerentes, que expressam uma relação mais verdadeira com a realidade sendo modelada. O especialista faz uso de metáforas de explicação para julgar características potencialmente problemáticas, ou até mesmo a ausência de características indispensáveis à modelagem.

[04] São frequentemente reportadas, em pesquisas de mercado, estatísticas que retratam a dificuldade em realizar projetos de aprendizado de máquina e inteligência artificial bem sucedidos. Acredita-se que uma das causas para a dificuldade em gerar valor a partir do aprendizado de máquina seja a típica abordagem de não incluir (de maneira adequada) o especialista no problema de interesse na

modelagem. Os especialistas (médicos, advogados, linguistas, artistas etc.) são geralmente envolvidos no julgamento do resultado, mas raramente participam do processo de modelagem. Uma possível justificativa para a falta de envolvimento do especialista no processo de modelagem é que os modelos produzidos são descritos como “caixa-preta”, ou seja, tipicamente só conseguimos ter acesso às predições do modelo. Dessa forma, os modelos ficam seriamente limitados ao conhecimento prévio do especialista, a seus vieses e suas crenças. Em suma, toda a promessa de modelagem pelo aprendizado de máquina fica comprometida quando o especialista não é envolvido nas decisões de modelagem.

[05] 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/>).

[06] Há no estado da técnica algumas tecnologias dedicadas a possibilitar a participação de humanos no processo de elaboração de

modelos baseados em inteligência artificial, como as apresentadas a seguir.

[07] O artigo científico intitulado “DALEX: Explainers for Complex Predictive Models in R”, de novembro de 2018, propõe uma metodologia que apresenta explicadores de modelos que são comparados entre si de modo a encontrar pontos fortes e fracos de diferentes modelos e fornece ferramentas adicionais para validação de modelo resultante (“*ensemble*”). A estrutura apresentada permite a compreensão da importância de uma determinada variável do modelo.

[08] 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 e 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.

[09] O documento de patente CN110709864 intitulado “HUMAN-IN-THE-LOOP INTERACTIVE MODEL TRAINING”, de 22/09/2018, apresenta um método para treinar um modelo preditivo que aumenta a interpretabilidade e confiabilidade do modelo para os usuários finais. Treina-se iterativamente o modelo, em cada ciclo de treinamento um humano poderá remover um ou mais predicados do modelo que não possuírem designações confiáveis, não forem causais ou ininteligíveis para a predição do modelo. O modelo final é então avaliado, por exemplo, quanto à precisão, complexidade, confiabilidade e interpretabilidade *post-hoc*.

[010] No estado da técnica não foram encontradas tecnologias com características semelhantes à tecnologia proposta neste pedido de patente, a saber: um processo centrado no humano para obtenção de

modelos baseados em Aprendizado de Máquina que inclui meios para permitir a um especialista (humano) contribuir para elaboração de modelos julgando-os, aplicando-se uma combinação de técnicas como a decomposição do espaço de dados em estruturas locais, do tipo “backbone”, formando um conjunto do tipo “aka backbone features”; representações vetoriais de explicações de modelos (explicabilidade); utiliza também agrupamento (“ensemble”) de modelos, dentre outras técnicas. As características técnicas do modelo proposto permitem alcançar as seguintes vantagens 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 contribuição de especialista humano ao modelo, utilizando metáforas de explicação para julgar características do modelo potencialmente problemáticas, ou até mesmo identificar a ausência de características indispensáveis à modelagem.

### **DESCRIÇÃO DETALHADA DA TECNOLOGIA**

[011] A tecnologia propõe um processo centrado no humano para obtenção de modelos baseados em aprendizado de máquina que inclui meios para permitir a um especialista (humano) contribuir para elaboração de modelos, julgando-os. A tecnologia utiliza técnicas como a decomposição do espaço de dados em estruturas locais, do tipo “backbone”, formando um conjunto do tipo “aka backbone features”; representações vetoriais de explicações de modelos (explicabilidade); agrupamento (“ensemble”) de modelos, dentre outras técnicas. A tecnologia viabiliza a contribuição de especialista humano ao modelo, utilizando metáforas de explicação para julgar características do modelo

potencialmente problemáticas, ou até mesmo identificar a ausência de características indispensáveis à modelagem.

[012] A tecnologia propõe alguns aspectos fundamentais: (i) a escolha de características para compor o modelo; e (ii) a explicação de como o modelo faz uso dessas características. Essas duas operações ocorrem concomitantemente, ou seja, a tecnologia realiza uma busca por características para compor um modelo e, durante a busca, alguns modelos são identificados como de melhor desempenho, o que implica na possibilidade de alto poder de predição. Para esses modelos de melhor desempenho, é realizada uma verificação de coerência, de forma a garantir um alinhamento maior entre a modelagem e a realidade. Diferentemente do que existe no estado da técnica, o objetivo do processo centrado no humano para elaboração de modelos baseados em aprendizado de máquina, ao final, é obter não apenas um modelo de aprendizado de máquina com alto poder de predição, mas também com alta coerência com a realidade sendo modelada. As etapas do processo proposto estão detalhadas a seguir.

### **1. Preparação**

[013] O primeiro passo envolve a criação de características. A título de exemplo, características podem ser concentrações de proteínas no plasma humano (no caso de modelagem para a criação de um exame de sangue), ou um conjunto de cubos sísmicos (no caso de modelagem para a predição de poços de petróleo). Nesse momento, é possível ter acesso a dados estruturados e tabulados (p.ex. tabelas), ou os dados podem ter uma forma não-estruturada (p.ex. texto, imagem, áudio, vídeo, grafos etc.). No caso de dados não-estruturados, é necessária uma etapa adicional que envolve a estruturação do dado, aqui denominada etapa de aprendizado de representação, já que, ao final, os dados se encontrarão representados de maneira tabulada. Na forma tabulada há um conjunto

de vetores, que são o único formato de entrada permitido para algoritmos de aprendizado de máquina.

## **2. Melhoria**

[014] O segundo passo é essencialmente uma repetição (um *loop*). O objetivo é que, ao final de cada iteração desse *loop*, haja uma interação com o especialista, daí o nome dado à tecnologia: “metodologia de aprendizado de máquina centrada no humano”. Aqui, um conjunto de modelos de melhor desempenho é apresentado ao especialista, que é capaz de realizar o julgamento do modelo através do uso de metáforas de explicação da predição. As metáforas podem passar diferentes tipos de informação ao especialista. Existem as metáforas para análise de populações. Por exemplo, tais metáforas podem informar a relação entre o valor aumentado (ou diminuído) da concentração de uma proteína e o quanto isso contribui para o diagnóstico de uma doença. Outra metáfora pode expor de que maneira diferentes características estão relacionadas. Há também as metáforas para análises individuais. Por exemplo, as metáforas que demonstram, para um paciente em particular, quais as proteínas mais impactam no diagnóstico realizado. O resultado dessa interação com o especialista é a criação ou adição de novas características, remoção de características aparentemente incoerentes, remoção de vieses nos dados, etc. Essa etapa de melhoria tem impactos teóricos e práticos. Do lado teórico, tipicamente o modelo é composto por um conjunto reduzido de características, o que leva a uma maior chance de generalização. Do outro lado, as características por vezes são avaliadas utilizando critérios práticos, como o custo para obtenção, ou sua confiabilidade.

## **3. Acionamento**

[015] Após uma série de iterações e, conseqüentemente, melhorias, o modelo é finalmente obtido e pode ser utilizado com maior segurança e

confiabilidade. Além disso, e de fato o que pode ser o fator mais importante, o modelo é naturalmente acionável, já que idealmente é coerente com a realidade. Isso implica em diversas consequências. Por exemplo, se o modelo permite prever quais proteínas (p.ex., características) estão mais relacionadas ao diagnóstico pretendido, o especialista pode criar teorias que resultem em novas drogas e terapias.

[016] Adiante apresentam-se as etapas técnicas necessárias à implementação do modelo.

[017] O processo centrado no humano para obtenção de modelos baseados em aprendizado de máquina compreende as seguintes etapas:

- a) Definir um espaço de dados como um conjunto de  $n$  pontos de dados da forma  $(x, y)$ , tal que  $x \in R_d$  e fornecer um vetor de característica  $\{x_1, x_2, \dots, x_d\}$ , onde  $y$  é a saída para uma entrada  $x$ , e os dados podem estar relacionados a rótulos;
- b) Decompor o espaço de dados da etapa “a” em subespaços contendo estruturas locais, do tipo “*backbone*”, de modo que haja um conjunto de características do tipo “*aka backbone features*”;
- c) Obter modelos por meio de treinamento a partir dos dados que representam uma combinação de subespaços individualmente relacionados a estruturas locais específicas definidas na etapa “b”;
- d) Realizar um processo de amostragem dos modelos obtidos na etapa “c” e selecionar, a partir da amostra obtida, os modelos de melhor desempenho de acordo com um critério baseado na medida de erro  $\ell(x')$  individual de cada modelo, procedendo a seleção pela comparação de um valor limiar de erro  $\epsilon$ , utilizando o critério de inclusão  $\ell(x') \leq \epsilon$ , onde forma-se um espaço  $(H')$  resultante de modelos contendo as possíveis explicações das predições relacionadas ao problema/fenômeno modelado;

- e) Gerar, para cada modelo no espaço ( $H'$ ) definido na etapa “d”, representações de suas preferências ( $p$ ) contidas em um vetor  $n$ -dimensional  $P(x') = \{p_1, p_2, \dots, p_n\}$ , 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  $e$ , de modo complementar, 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  $E(x') = \{e_1, e_2, \dots, e_d\}$ , onde  $e_i$  corresponde à influência que a característica  $x_i$  exerce na predição realizada pelo modelo  $x'$ ;
- g) Agrupar os modelos do espaço ( $H'$ ) em *clusters* (ou 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 o modelo de melhor desempenho dentro do conjunto de modelos que forma cada um dos clusters/aglomerados envolvendo um especialista humano na seleção e utilizando os fatores explicativos e metáforas de explicação;

- i) Formar um agrupamento (*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 do agrupamento, o rótulo (que está associado ao seu respectivo modelo) que obtiver a maior quantidade de votos.

[018] Acrescentam-se algumas técnicas que podem ser utilizadas na concretização da presente tecnologia, como na etapa “d”, onde as características que compõem cada modelo  $x'$  são selecionadas aleatoriamente, sendo possível utilizar árvores com aumento de gradiente e a técnica “*random forests*” como algoritmos de aprendizagem, dentre outras técnicas.

[019] A utilização dos valores “*SHapley Additive exPlanations*” (SHAP) permite avaliar a importância das características (“*features*”) na predição do modelo e pode ser utilizada na etapa “f”.

[020] Na etapa “g”, o critério de formação dos aglomerados pode ser baseado na distância entre seus vetores de explicação e, para maximizar a coesão e separação dos *clusters*, pode-se utilizar um critério com base na distância envolvendo a preferência de modelo.

[021] A presente tecnologia pode ser usada, de forma não limitante, para as seguintes aplicações: (i) predição da evolução do quadro clínico e tempo de permanência de pacientes internados em Unidades de Tratamento Intensivos (UTI), através do monitoramento de análises laboratoriais e de sinais vitais, de glicose e pressão arterial (PA), até níveis de adrenalina e  $VO_2$ ; (ii) no diagnóstico de COVID-19 através dos parâmetros do hemograma de sangue; (iii) na predição de Alzheimer na fase pré-sintomática através do proteoma plasmático.

[022] Para o uso (i), a tecnologia pode apresentar as seguintes etapas: a) capturar as interações fisiológicas locais, como a frequência cardíaca, as medidas de creatinina, a pressão arterial sistólica no nível mais baixo, usando uma Rede Neural Convolutacional (CNN) para extrair recursos de vários níveis abstratos; b) extrair dependências de longo alcance baseadas em sinais fisiológicos no nível superior usando uma long short-memory (LSTM) explorando informações temporais dentro de sinais vitais e dados laboratoriais para dinamicamente prever os resultados do paciente e usar uma sequência desses recursos para gerar previsões dinâmicas para a mortalidade do paciente; c) delimitar espaços de risco de mortalidade e avaliar a proximidade de regiões de risco e a velocidade com que a condição do paciente muda.

[023] Para o uso (ii), a tecnologia pode apresentar as seguintes etapas: a) extração de hemogramas coletados na janela de 24 horas do primeiro resultado RT-PCR positivo para COVID (COVID Ativo); b) extração de hemogramas coletados antes de 2020 e associados a um resultado positivo para outros vírus ou patologias confundidoras; c) extração de hemogramas coletados na janela de 24 horas de um resultado RT-PCR negativo para COVID e explicitamente anterior ao primeiro RT-PCR positivo de um dado paciente (Pré-COVID); d) filtragem de hemogramas com valores de analitos muito destoantes do padrão (mais de dois desvios dos valores de referência) para remoção de ruído; e) aprender padrões sanguíneos tipicamente associados com quadros de COVID-19 e não COVID-19. São possíveis os seguintes tratamentos para o treinamento dos modelos: a) criar de múltiplos modelos especializados e associados a um resultado positivo para outros vírus de RNA ou patologias confundidoras; b) combinar as saídas dos modelos obtidos na etapa “a” para obter uma predição final acerca da doença alvo; c) treinar um modelo de aprendizado de máquina baseado em árvores de decisão

com “*gradient boosting*” e utilizar o pacote SHAP para obter uma interpretação da predição do modelo, de forma a criar uma ferramenta de apoio à decisão médica. Pode-se ainda criar um modelo único que leve em consideração vírus de RNA ou patologias confundidoras incluídas nos dados.

[024] Para o uso (iii) a tecnologia pode compreender um processo de explicação do modelo por atribuição de importância dos biomarcadores associados à conversão de indivíduos com comprometimento cognitivo leve (MCI) em Doença de Alzheimer (AD) para explicação da progressão da doença e exploração e compreensão dos fatores que mais contribuem para as decisões do modelo, compreendendo as seguintes etapas: a) calcular a importância de um biomarcador dentro de um modelo, estimando o efeito de sua ausência na decisão do modelo usando o algoritmo SHAP para cada indivíduo, sendo que o efeito da ausência do biomarcador é calculado pela obtenção da distância da previsão original; b) se a distância for negativa, o novo valor previsto é maior para a conversão AD do que a previsão original, assim ao indivíduo em questão atribui-se um valor de predição que denota um risco reduzido de conversão em comparação com o risco médio para a população; c) calcular uma média de distâncias absolutas entre os indivíduos e gerar uma lista classificada de biomarcadores de acordo com o efeito nas previsões de resultados em todos os indivíduos; d) as interações de biomarcadores individuais são calculadas para os analitos de melhor classificação; e) o biomarcador em questão (analito de melhor classificação) é então removido do modelo, e cada um dos biomarcadores restantes são removidos iterativamente para estimar o efeito de ambos os biomarcadores estarem ausentes ao mesmo tempo; f) o efeito da interação é comparado com o efeito aditivo dos biomarcadores individualmente; g) a diferença entre a soma dos efeitos

ao remover os dois biomarcadores individualmente e o efeito da remoção dos dois biomarcadores simultaneamente é calculada para cada indivíduo e a média das distâncias absolutas é usada para encontrar os principais atributos de interação que são utilizados para explicação.

[025] A presente tecnologia pode ser mais bem compreendida através dos exemplos abaixo, não limitantes.

### **EXEMPLO 1 – TECNOLOGIA APLICADA PARA SUPORTE À DECISÃO MÉDICA**

[026] Apresenta-se uma possibilidade de concretização da tecnologia que consiste em uma solução integrada de predição para melhorar a qualidade das Unidades de Tratamento Intensivos, quanto à tomada de decisão a respeito de um paciente internado. Essa solução se baseia em um modelo de Inteligência artificial (IA) (ligado à tecnologia da Metodologia de Aprendizado de Máquina Centrada no Humano) capaz de transformar dados dos pacientes em informações que auxiliem o intensivista a gerir melhor aqueles de alto-risco e alto-custo. A tecnologia apresenta, como efeito surpreendente, a aplicação para a predição da evolução do quadro clínico e tempo de permanência de pacientes internados em UTI, através do monitoramento de análises laboratoriais e de sinais vitais (de glicose e pressão arterial (PA), até níveis de adrenalina, VO<sub>2</sub>, etc).

[027] Soluções preditivas em UTIs são de imensa importância na gestão hospitalar, com impacto na qualidade do atendimento ao paciente e na otimização de recursos hospitalares. Nos EUA, são gastos U\$ 82 bilhões/ano com internações em UTIs, correspondendo a 13% dos custos hospitalares e 4% dos custos totais com saúde (dados do *Society of Critical Care Medicine*). No Brasil, apenas o SUS tem gastos de aproximadamente R\$ 3,0 bilhões/ano com internações em UTIs (dados do Datasus). Tamanho investimento se justifica, pois a UTI é etapa

fundamental no acolhimento de pacientes em estado grave, visando o aumento das chances de sobrevivência. Devido à importância da área, múltiplas pesquisas já foram conduzidas para predição de resultados em pacientes na UTI. Entretanto, a precisão dos modelos gerados é boa apenas para grupos de pacientes (i.e., SAPS-3), sendo extremamente fraca para pacientes individualmente.

[028] As Unidades de Terapia Intensiva (UTI) são locais de atendimento de pacientes graves ou de risco e recuperáveis que dispõem de assistência médica ininterrupta, contínua, com equipamentos específicos próprios, recursos humanos especializados e que tenham acesso a tecnologias destinadas a diagnóstico e terapêutica. A quantidade de dados gerados por segundo em uma unidade de UTI é tão grande que não é possível ser assimilada por uma pessoa em tempo real, pois são de complexo nível de entendimento e gerados em ritmo muito maior do que os médicos intensivistas podem absorver. Modelos de Inteligência Artificial vêm sendo testados para, através de uma entrada de um banco extenso de dados da UTI, auxiliarem os médicos elencando os riscos atuais e futuros de um paciente bem como os recursos a serem utilizados em sua permanência. O intensivista e sua equipe possuem uma quantidade extensa de possibilidades e recursos que precisam ser gerenciados e o sistema de UTI Inteligente entra como ferramenta fundamental de auxílio no processo de tomada de decisão. Hoje, a avaliação individual de risco clínico (complicações e permanência) de pacientes durante o atendimento hospitalar é bastante difícil e há problemas no acompanhamento e predição da jornada do paciente desde sua entrada no hospital até momento da alta médica.

[029] A presente tecnologia tem como base técnicas de Inteligência Artificial, que simulam, por meio de algoritmos específicos, mecanismos avançados de cognição e suporte à decisão baseados em grandes

volumes de informação. Seu funcionamento alicerça-se no uso de aprendizado de máquina aplicada ao suporte à decisão do humano. As características são escolhidas para a composição do modelo e através de técnicas de aprendizado de máquina são feitas as previsões relativas à internação do paciente. O modelo apresenta também a explicação de como o modelo faz uso dessas características, o que pode se tornar informação clínica valiosa. Todo o desenvolvimento tem como intuito aumentar a capacidade do intensivista de identificar e gerir pacientes de alto-risco e alto-custo.

[030] Os dados dos pacientes na Unidade de Terapia Intensiva (UTI) são extensos, complexos e frequentemente produzidos a uma taxa muito maior do que os intensivistas podem absorver. O monitoramento de pacientes de UTI está se tornando cada vez mais desafiador, e os sistemas que aprendem com os dados da UTI para alertar os médicos sobre os riscos atuais e futuros de um paciente estão desempenhando um papel significativo no processo de tomada de decisão.

[031] Neste exemplo é explorada a abordagem de adaptação de domínio para melhorar a acurácia dos sistemas de avaliação que apresentam condições de treino e teste incompatíveis. São propostos modelos profundos que extraem o domínio compartilhado e as características latentes específicas de domínios de UTI ou subpopulações de pacientes. Cada domínio corresponde a um tipo diferente de UTI, como UTI cardíaca, coronariana, médica e cirúrgica. Isso nos permite aprender vários modelos que são específicos para cada domínio da UTI, melhorando a precisão da previsão em relação às populações de pacientes. Para isso, são empregadas abordagens de transferência que se diferem em termos da escolha de quais camadas congelar ou ajustar. São propostos modelos de aprendizado profundo que combinam redes convolucionais e recorrentes. Além da previsão de mortalidade,

capturamos as interações fisiológicas locais (por exemplo, frequência cardíaca, creatinina, pressão arterial sistólica) no nível mais baixo usando uma Rede Neural Convolutiva (CNN) e extraímos dependências de longo alcance baseadas em sinais fisiológicos no nível superior usando uma rede de memória de longo prazo (LSTM). Assim, os modelos exploram informações temporais dentro de sinais vitais e dados laboratoriais para dinamicamente prever os resultados do paciente, ou seja, o componente CNN extrai recursos de vários níveis abstratos e o componente LSTM ingere uma sequência desses recursos para gerar previsões dinâmicas para a mortalidade do paciente. Com o uso destes modelos foram obtidos valores de área sob a curva (AUC) superiores a 83.

[032] Como consequência, as representações aprendidas junto com as previsões para um paciente específico durante a permanência na UTI formam a trajetória do paciente correspondente e, assim, um espaço de risco de mortalidade pode ser obtido a partir de um conjunto de trajetórias anteriores do paciente. O benefício fundamental de analisar as trajetórias futuras dos pacientes no espaço de risco de mortalidade é o foco na dinâmica, enfatizando a proximidade de regiões de risco e a velocidade com que a condição do paciente muda. Assim, o espaço de risco de mortalidade permite aos médicos rastrear tendências de risco e obter informações sobre suas decisões de tratamento.

[033] Apresenta-se o uso do processo centrado no humano para elaboração de modelos baseados em aprendizado de máquina para predição da evolução do quadro clínico e tempo de permanência de pacientes internados em Unidades de Tratamento Intensivos (UTI) através do monitoramento de análises laboratoriais e de sinais vitais (de glicose e pressão arterial (PA), até níveis de adrenalina,  $VO_2$ ). O processo compreende as seguintes etapas:

- a) capturar as interações fisiológicas locais, por exemplo, a frequência cardíaca, as medidas de creatinina, a pressão arterial sistólica no nível mais baixo usando uma Rede Neural Convolucional (CNN) para extrair recursos de vários níveis abstratos;
- b) extrair dependências de longo alcance baseadas em sinais fisiológicos no nível superior usando uma long short-memory (LSTM) explorando informações temporais dentro de sinais vitais e dados laboratoriais para dinamicamente prever os resultados do paciente e usar uma sequência desses recursos para gerar previsões dinâmicas para a mortalidade do paciente;
- c) delimitar espaços de risco de mortalidade e avaliar a proximidade de regiões de risco e a velocidade com que a condição do paciente muda.

## **EXEMPLO 2 – TECNOLOGIA APLICADA PARA DIAGNÓSTICO COVID-19 ATRAVÉS DO HEMOGRAMA DE SANGUE**

[034] Este exemplo representa uma solução para auxiliar na triagem de diagnósticos de pacientes infectados com o SARS-Cov-2 (relacionada à Metodologia de Aprendizado de Máquina centrada no humano). Para isso foi desenvolvido um modelo de inteligência artificial que prevê o resultado do RT-PCR (do inglês: “*Reverse transcription polymerase chain Reaction*”), considerado padrão ouro para diagnóstico da Covid-19, usando apenas analitos de um hemograma completo.

[035] Para um efetivo controle da pandemia, são necessárias informações confiáveis e cotidianamente atualizadas de forma que possam então auxiliar políticas públicas na tomada de decisão a fim de desacelerar o curso pandêmico. Dessa forma, é essencial que órgãos governamentais tenham acesso ao número de pacientes atualmente infectados que precisam ficar em quarentena e também o número de novos infectados nas últimas 24 horas. Por isso, um método eficaz, rápido e acessível torna-se fundamental.

[036] Atualmente, o teste de RT-PCR não está amplamente disponível ou acessível na maioria dos países em desenvolvimento, incluindo o Brasil, o que pode aumentar as taxas de contágio e atrasar medidas preventivas importantes. Assim, identificar / avaliar o risco de pacientes serem infectados usando apenas testes baratos e rotineiramente disponíveis, como o CBC, para o qual a maioria dos países já possui toda a infraestrutura necessária instalada, poderia auxiliar na tomada de decisão clínica imediata, no uso efetivo de recursos para a saúde e também influenciar em políticas públicas.

[037] A tecnologia tem como base técnicas de Inteligência Artificial, que simulam, por meio de algoritmos específicos, mecanismos avançados de cognição e suporte à decisão baseados em grandes volumes de informação. Seu funcionamento alicerça-se no uso de aprendizado de máquina aplicada ao suporte à decisão do humano. Um hemograma completo é um conjunto de testes de laboratório que apresentam dados sobre as células sanguíneas. Os exames de hemograma indicam a quantidade de glóbulos brancos, glóbulos vermelhos, plaquetas e a concentração de hemoglobina e hematócrito. Usando esses dados do hemograma, foram aplicados modelos de inteligência artificial que combinam os analitos para detectar se o paciente tem ou não COVID-19, sendo que a predição é realizada através de técnica de explicabilidade (Pesquisa atrelada a tecnologia Metodologia de Aprendizado de Máquina centrada no humano). Esse modelo utiliza algoritmos de aprendizado de máquina supervisionado, ou seja, ele é construído usando um banco de dados rotulado.

[038] Por fazerem parte de baterias de exames distintos, frequentemente hemogramas e PCRs são coletados em datas distintas. Uma das principais perguntas de pesquisa que buscamos responder foi estudar qual seria a janela de tempo ideal entre um hemograma e um PCR para

validar seu rótulo. Suponha que um hemograma seja coletado no dia 1 e obtenhamos um resultado RT-PCR positivo no dia 5. É correto afirmar que este hemograma está associado com uma pessoa doente? Sem informações sobre o início dos sintomas, não é possível validar se esse hemograma realmente caracteriza um paciente doente. O mesmo ocorre em hemogramas coletados após a data do primeiro resultado positivo para PCR. Durante nossa análise investigatória averiguamos que o padrão sanguíneo dos pacientes se torna alterado após o início de tratamentos. Após estudos com diferentes períodos, concluímos que o momento mais efetivo para coleta dos dados é a janela de 24 horas centrada no primeiro resultado PCR positivo de um paciente, e o restante é altamente incerto em relação a um diagnóstico positivo.

[039] Outro ponto de atenção é a alta taxa de resultados falso-negativos para o RT-PCR. Em particular, observamos frequentemente casos de um mesmo paciente tendo resultados negativos intercalados por dois ou mais resultados positivos. Assim sendo, também é necessário realizar um tratamento de dados para garantir a veracidade dos rótulos negativos e garantir que o modelo seja o mais fiel possível ao cenário real do COVID, e não às limitações do exame RT-PCR. Outro fator relevante é a baixa prevalência de outras doenças infecciosas no período de 2020. Devido aos hábitos adotados pela população, foi observada uma baixa quantidade de casos de influenza e outras doenças com quadros sintomáticos semelhantes ao de COVID. Dessa forma, uma preocupação extra é garantir que o modelo seja de fato capaz de diferenciar um paciente com COVID de um sem, e não simplesmente um paciente saudável de um doente.

[040] Podemos sumarizar as etapas de pré-processamento nos seguintes passos:

- a) Extração de hemogramas coletados na janela de 24 horas do primeiro resultado RT-PCR positivo para COVID (COVID Ativo);
- b) Extração de hemogramas coletados antes de 2020 e associados a um resultado RT-PCR positivo para Influenza-A, Influenza-B, Influenza-H1N1, Influenza Sazonal ou Virusmol (Influenza Ativo);
- c) Extração de hemogramas coletados na janela de 24 horas de um resultado RT-PCR negativo para COVID e explicitamente anterior ao primeiro RT-PCR positivo de um dado paciente (Pré-COVID);
- d) Filtragem de hemogramas com valores de analitos (parâmetros do hemograma: células sanguíneas) muito destoantes do padrão (mais de dois desvios dos valores de referência) para remoção de ruído.

[041] A etapa de treinamento consiste na criação de múltiplos modelos especializados em cada uma das doenças avaliadas, e combinar suas saídas para obter uma predição final acerca da doença alvo. Uma segunda possibilidade é, uma vez estabelecido o conjunto limpo de dados, treinamos um modelo de aprendizado de máquina baseado em árvores de decisão com *gradient boosting* (lightGBM, <https://lightgbm.readthedocs.io/en/latest/>) e utilizamos o pacote SHAP (<https://shap.readthedocs.io/en/latest/>) para obter uma interpretação da predição do modelo, de forma a criar uma ferramenta que seja útil no apoio à decisão médica. Isso nos permite ter não somente uma predição acerca do COVID, mas também a explicação do resultado obtido (ie. acredita-se que o paciente não tenha COVID, pois seus leucócitos estão acima do padrão para COVID, etc.). Avaliamos os resultados obtidos tanto por meio de validação cruzada com 5 “*folds in-sample*”, quanto

separando um período dos dados para teste “*out-of-sample*”, obtendo valores de área sob a curva (AUC) superiores a 90 em ambos os casos.

[042] Apresenta-se o uso do processo centrado no humano para elaboração de modelos baseados em aprendizado de máquina para diagnóstico de COVID-19 através dos parâmetros do hemograma de sangue, que compreende os seguintes tratamentos para o pré-processamento de dados:

- a) extração de hemogramas coletados na janela de 24 horas do primeiro resultado RT-PCR positivo para COVID (COVID Ativo);
- b) extração de hemogramas coletados antes de 2020 e associados a um resultado positivo para outros vírus ou patologias confundidoras;
- c) extração de hemogramas coletados na janela de 24 horas de um resultado RT-PCR negativo para COVID e explicitamente anterior ao primeiro RT-PCR positivo de um dado paciente (Pré-COVID);
- d) filtragem de hemogramas com valores de analitos muito destoantes do padrão (mais de dois desvios dos valores de referência) para remoção de ruído;
- e) aprender padrões sanguíneos tipicamente associados com quadros de COVID-19 e não COVID-19.

[043] O processo também inclui os seguintes tratamentos para o treinamento dos modelos:

- a) criar de múltiplos modelos especializados e associados a um resultado positivo para outros vírus de RNA ou patologias confundidoras;
- b) combinar as saídas dos modelos obtidos na etapa “a” para obter uma predição final acerca da doença alvo;
- c) treinar um modelo de aprendizado de máquina baseado em árvores de decisão com “*gradient boosting*” e utilizar o pacote

SHAP para obter uma interpretação da predição do modelo, de forma a criar uma ferramenta de apoio à decisão médica.

[044] O processo permite, de forma alternativa, a criação de um modelo único que leve em consideração vírus de RNA ou patologias confundidoras incluídas nos dados.

### **EXEMPLO 3 – TECNOLOGIA APLICADA PARA PREDIÇÃO DE ALZHEIMER NA FASE PRÉ-SINTOMÁTICA ATRAVÉS DO PROTEOMA PLASMÁTICO**

[045] Este exemplo representa uma solução a partir do uso de inteligência artificial para desenvolver imunoenaios capazes de diagnosticar a doença de Alzheimer, ainda na fase pré-sintomática, quando existem mecanismos de retardo ou prevenção da doença.

[046] A solução se baseia em um modelo de Inteligência artificial (IA), ligado à tecnologia da Metodologia de Aprendizado de Máquina Centrada no Humano.

[047] O número atual de pacientes com Alzheimer é de 47 milhões, com estimativas para aumentar para 150 milhões até 2050, sendo considerada a doença mais temida do mundo. Os exames para diagnosticar a doença são caros e invasivos, envolvendo punção lombar e exames com alta radioatividade. Por isso, esses exames são realizados apenas quando os sintomas já estão evidentes.

[048] O prognóstico do paciente é sempre muito ruim. Embora existam tratamentos disponíveis que possam melhorar alguns sintomas da doença, não há cura ou terapia modificadora disponível atualmente, e a doença inevitavelmente progride em todos os pacientes. Avanços começam a apontar para possíveis drogas (ainda necessitando de aprovação), como o aducanumab, que atua como um redutor de amilóide.

[049] Muito provavelmente, as primeiras drogas para a Doença de Alzheimer deverão ser analisadas sob duas perspectivas. Primeiramente, tais drogas virão com efeitos adversos. De outro lado, a eficácia desses medicamentos é potencializada caso sejam aplicadas precocemente. Portanto, torna-se de primordial importância o diagnóstico antecipado da Doença de Alzheimer, ainda na fase pré-sintomática. É nesse contexto que a presente tecnologia ganha relevância, como um exame não-invasivo, seguro e de alta assertividade, possibilitando ao paciente seguir o tratamento apenas quando este for realmente necessário.

(<https://sbgg.org.br/com-o-envelhecimento-populacional-doenca-de-alzheimer-devera-aumentar-nas-proximas-decadas-aponta-sbgg/#:~:text=Segundo%20estimativas%20da%20Alzheimer's%20Disease,pessoas%20em%20todo%20o%20mundo>),

([https://www.ft.com/content/6a648aaf-209d-4186-be29-](https://www.ft.com/content/6a648aaf-209d-4186-be29-31f456d91b78Sevigny)

31f456d91b78Sevigny), (J, Chiao P, Bussiere T, et al. The antibody aducanumab reduces A $\beta$  plaques in Nature 2016; 537(7618):50–56).

[050] A tecnologia tem como base técnicas de Inteligência Artificial que simulam, por meio de algoritmos específicos, mecanismos avançados de cognição e suporte à decisão baseados em grandes volumes de informação. Seu funcionamento alicerça-se no uso de aprendizado de máquina aplicado ao suporte à decisão do humano detectando pacientes com alto risco de desenvolverem os sintomas neurodegenerativos da doença de Alzheimer, a partir da análise da proteômica plasmática. É feita a identificação dos pacientes com maior probabilidade de converterem para a doença de Alzheimer em um horizonte de até 4 anos. Além disso, a ferramenta é capaz de identificar quais os mecanismos biológicos envolvidos no avanço da doença.

[051] Processos biológicos complexos associados à progressão da Doença de Alzheimer (DA) são mediados por interações entre diversas

proteínas. Uma vez que essas interações não óbvias não são totalmente compreendidas, desenvolvemos uma exploração em larga escala de modelos de risco de desempenho, cada um explorando diferentes interações entre os biomarcadores. Especificamente, amostramos mais de um bilhão de modelos para um horizonte de conversão de 4 anos, onde cada modelo é composto por uma combinação diferente de biomarcadores. Os biomarcadores dentro de um modelo são selecionados aleatoriamente, resultando em modelos com desempenho preditivo diversificado.

[052] Os modelos são construídos usando uma implementação rápida do algoritmo LightGBM, que produz um modelo complexo composto por centenas de árvores de decisão simples que são finalmente combinadas em um único modelo por um processo conhecido como *boosting*. Essa complexidade adicional permite que esses modelos alcancem altos níveis de precisão que não podem ser obtidos por modelos simples. O desempenho preditivo de cada modelo LightGBM é avaliado em termos da área padrão sob a medida da curva ROC (AUROC).

[053] Avaliamos mais de um bilhão de modelos LightGBM. Para cada modelo, conduzimos cinquenta rodadas de validação cruzada quádrupla, melhorando assim a robustez das estimativas e permitindo controlar a estabilidade dos resultados. Os dados ADNI são organizados em cinco grupos. Em cada rodada, quatro grupos são usados como um conjunto de treinamento e o quinto grupo é usado como um conjunto de teste. Todo o procedimento quádruplo é repetido cinquenta vezes e, para cada modelo, relatamos a média de suas médias AUROC. Os números reportados são dados em termos de médias de médias e, portanto, o erro padrão da análise de médias garante a confiabilidade dos modelos selecionados. Com o uso destes modelos, foram obtidos valores de área sob a curva (AUC) superiores a 88.

[054] Explicação do modelo por atribuição de importância do biomarcador: Existem muitos estudos que analisam biomarcadores significativos associados à conversão de indivíduos com comprometimento cognitivo leve (MCI) em AD, mas os resultados têm sido inconsistentes. Diferentes biomarcadores foram avaliados, no entanto, MCI é um grupo diverso, e a progressão de MCI para AD pode ocorrer devido a uma infinidade de razões. Portanto, é improvável que uma única proteína preveja todas as conversões possíveis. Em contraste, combinamos diferentes proteínas em um modelo único e de alto desempenho, de modo que diversos mecanismos de conversão potenciais podem ser levados em consideração. Este modelo de alto desempenho pode ser visto como uma explicação possível (mas talvez incompleta) para a progressão da doença, uma vez que os fatores que levam à decisão do modelo são baseados nos biomarcadores dentro do modelo. Assim, queremos permitir a formação de insights, explorando e compreendendo os fatores que mais contribuem para as decisões do modelo.

[055] Calculamos a importância de um biomarcador dentro de um modelo estimando o efeito de sua ausência na decisão do modelo usando o algoritmo SHAP. Para cada indivíduo, o efeito da ausência do biomarcador é calculado pela obtenção da distância da previsão original. Uma distância negativa é obtida se o novo valor previsto for maior para a conversão AD do que a previsão original, sugerindo que o indivíduo em questão tem um valor que reduziu o risco de conversão em comparação com o risco médio para a população.

[056] Uma média de distâncias absolutas entre os indivíduos é usada para gerar uma lista classificada de biomarcadores de acordo com o efeito nas previsões de resultados em todos os indivíduos. As interações de biomarcadores individuais são calculadas para os analitos de melhor

classificação. Novamente, o biomarcador em questão é removido do modelo, e cada um dos biomarcadores restantes são removidos iterativamente para estimar o efeito de ambos os biomarcadores estarem ausentes ao mesmo tempo. O efeito da interação é comparado com o efeito aditivo dos biomarcadores individualmente. A diferença entre a soma dos efeitos ao remover os dois biomarcadores individualmente e o efeito da remoção dos dois biomarcadores simultaneamente é calculada para cada indivíduo e a média das distâncias absolutas é usada para encontrar os principais atributos de interação. (Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiewi Ye, Tie-Yan Liu: LightGBM: A highly efficient gradient boosting decision tree. NIPS. 2017; 3146 - 3154.); (Yoav Freund and Robert Schapire: A decision-theoretic generalization of online learning and an application to boosting, Journal of Computer and System Sciences. 1997; 55(1): 119 – 139); (Tom Fawcett: An introduction to ROC analysis. Pattern Recognition Letters. 2006; 27(8): 861 - 874.); (Scott Lundberg and Su-In Lee: A unified approach to interpreting model predictions. NIPS. 2017: 4765 – 4774).

[057] Apresenta-se o uso do processo centrado no humano para elaboração de modelos baseados em aprendizado de máquina para predição de Alzheimer na fase pré-sintomática através do proteoma plasmático. Trata-se de um processo de explicação do modelo por atribuição de importância dos biomarcadores associados à conversão de indivíduos com comprometimento cognitivo leve (MCI) em Doença de Alzheimer (AD) para explicação da progressão da doença, exploração e compreensão dos fatores que mais contribuem para as decisões do modelo. O processo compreende as seguintes etapas:

- a) calcular a importância de um biomarcador dentro de um modelo estimando o efeito de sua ausência na decisão do modelo

usando o algoritmo SHAP para cada indivíduo, sendo que o efeito da ausência do biomarcador é calculado pela obtenção da distância da previsão original;

b) se a distância for negativa, o novo valor previsto é maior para a conversão AD do que a previsão original, assim ao indivíduo em questão atribui-se um valor de predição que denota um risco reduzido de conversão em comparação com o risco médio para a população;

c) calcular uma média de distâncias absolutas entre os indivíduos e gerar uma lista classificada de biomarcadores de acordo com o efeito nas previsões de resultados em todos os indivíduos;

d) as interações de biomarcadores individuais são calculadas para os analitos de melhor classificação;

e) o biomarcador em questão (analito de melhor classificação) é então removido do modelo, e cada um dos biomarcadores restantes é removido iterativamente para estimar o efeito de ambos os biomarcadores estarem ausentes ao mesmo tempo;

f) o efeito da interação é comparado com o efeito aditivo dos biomarcadores individualmente;

g) a diferença entre a soma dos efeitos ao remover os dois biomarcadores individualmente e o efeito da remoção dos dois biomarcadores simultaneamente é calculada para cada indivíduo e a média das distâncias absolutas é usada para encontrar os principais atributos de interação que são utilizados para explicação.

## REIVINDICAÇÕES

### 1. PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA **caracterizado por** compreender as seguintes etapas:

- a) Definir um espaço de dados como um conjunto de  $n$  pontos de dados da forma  $(x, y)$ , tal que  $x \in R_d$  e fornecer um vetor de característica  $\{x_1, x_2, \dots, x_d\}$ , onde  $y$  é a saída para uma entrada  $x$ , e os dados podem estar relacionados a rótulos;
- b) Decompor o espaço de dados da etapa “a” em subespaços contendo estruturas locais, do tipo “*backbone*”, de modo que haja um conjunto de características do tipo “*aka backbone features*”;
- c) Obter modelos por meio de treinamento a partir dos dados que representam uma combinação de subespaços individualmente relacionados a estruturas locais específicas definidas na etapa “b”;
- d) Realizar um processo de amostragem dos modelos obtidos na etapa “c” e selecionar, a partir da amostra obtida, os modelos de melhor desempenho de acordo com um critério baseado na medida de erro  $\ell(x')$  individual de cada modelo, procedendo a seleção pela comparação de um valor limiar de erro  $\epsilon$ , utilizando o critério de inclusão  $\ell(x') \leq \epsilon$ , onde forma-se um espaço  $(H')$  resultante de modelos contendo as possíveis explicações das predições relacionadas ao problema/fenômeno modelado;
- e) Gerar, para cada modelo no espaço  $(H')$  definido na etapa “d”, representações de suas preferências  $(p)$  contidas em um vetor  $n$ -dimensional  $P(x') = \{p_1, p_2, \dots, p_n\}$ , 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 e, de modo complementar, 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  $E(x') = \{e_1, e_2, \dots, e_d\}$ , onde  $e_i$  corresponde à influência que a característica  $x_i$  exerce na predição realizada pelo modelo  $x'$ ;

g) Agrupar os modelos do espaço ( $H'$ ) em *clusters* (ou 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 o modelo de melhor desempenho dentro do conjunto de modelos que forma cada um dos clusters/aglomerados envolvendo um especialista humano na seleção e utilizando os fatores explicativos e metáforas de explicação;

i) Formar um agrupamento (*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 do agrupamento, o rótulo (que está associado ao seu respectivo modelo) que obtiver a maior quantidade de votos.

**2. PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA, de acordo com a reivindicação 1 caracterizado por, na etapa “d”, utilizar árvores**

com aumento de gradiente e a técnica “*random forests*” como algoritmos de aprendizagem.

**3. PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA, de acordo com a reivindicação 1 caracterizado por, na etapa “f”, utilizar valores “SHapley Additive exPlanations” (SHAP) para avaliar a importância das características (“*features*”) na predição do modelo.**

**4. PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA, de acordo com a reivindicação 1 caracterizado por, na etapa “g”, o critério de formação dos aglomerados ser baseado na 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.**

**5. USO DO PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA definido na reivindicação 1, caracterizado por ser para predição da evolução do quadro clínico e tempo de permanência de pacientes internados em Unidades de Tratamento Intensivos (UTI), através do monitoramento de análises laboratoriais e de sinais vitais, de glicose e pressão arterial (PA), até níveis de adrenalina e VO<sub>2</sub>.**

**6. USO DO PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA, de acordo com a reivindicação 5, caracterizado pela predição da evolução do quadro clínico e tempo de permanência de pacientes internados em Unidades de Tratamento Intensivos (UTI) e compreender as seguintes etapas:**

a) capturar as interações fisiológicas locais, como a frequência cardíaca, as medidas de creatinina, a pressão arterial sistólica no nível mais baixo,

usando uma Rede Neural Convolucional (CNN) para extrair recursos de vários níveis abstratos;

b) extrair dependências de longo alcance baseadas em sinais fisiológicos no nível superior usando uma *long short-memory* (LSTM) explorando informações temporais dentro de sinais vitais e dados laboratoriais para dinamicamente prever os resultados do paciente e usar uma sequência desses recursos para gerar previsões dinâmicas para a mortalidade do paciente;

c) delimitar espaços de risco de mortalidade e avaliar a proximidade de regiões de risco e a velocidade com que a condição do paciente muda.

**7. USO DO PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA** definido na reivindicação 1, caracterizado por ser no diagnóstico de COVID-19 através dos parâmetros do hemograma de sangue.

**8. USO DO PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA**, de acordo com a reivindicação 7, caracterizado por compreender os seguintes tratamentos para o pré-processamento de dados:

a) extração de hemogramas coletados na janela de 24 horas do primeiro resultado RT-PCR positivo para COVID (COVID Ativo);

b) extração de hemogramas coletados antes de 2020 e associados a um resultado positivo para outros vírus ou patologias confundidoras;

c) extração de hemogramas coletados na janela de 24 horas de um resultado RT-PCR negativo para COVID e explicitamente anterior ao primeiro RT-PCR positivo de um dado paciente (Pré-COVID);

d) filtragem de hemogramas com valores de analitos muito destoantes do padrão (mais de dois desvios dos valores de referência) para remoção de ruído;

e) aprender padrões sanguíneos tipicamente associados com quadros de COVID-19 e não COVID-19.

**9. USO DO PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA, de acordo com a reivindicação 7, caracterizado por compreender os seguintes tratamentos para o treinamento dos modelos:**

a) criar de múltiplos modelos especializados e associados a um resultado positivo para outros vírus de RNA ou patologias confundidoras;

b) combinar as saídas dos modelos obtidos na etapa “a” para obter uma predição final acerca da doença alvo;

c) treinar um modelo de aprendizado de máquina baseado em árvores de decisão com “*gradient boosting*” e utilizar o pacote SHAP para obter uma interpretação da predição do modelo, de forma a criar uma ferramenta de apoio à decisão médica.

**10. USO DO PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA, de acordo com a reivindicação 9, caracterizado pela etapa “a” compreender a criação de um modelo único que leve em consideração vírus de RNA ou patologias confundidoras incluídas nos dados.**

**11. USO DO PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA definido na reivindicação 1, caracterizado por ser na predição de Alzheimer na fase pré-sintomática através do proteoma plasmático.**

**12. USO DO PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA, de acordo com a reivindicação 11, caracterizado por**

compreender um processo de explicação do modelo por atribuição de importância dos biomarcadores associados à conversão de indivíduos com comprometimento cognitivo leve (MCI) em Doença de Alzheimer (AD) para explicação da progressão da doença e exploração e compreensão dos fatores que mais contribuem para as decisões do modelo, compreendendo as seguintes etapas:

- a) calcular a importância de um biomarcador dentro de um modelo, estimando o efeito de sua ausência na decisão do modelo usando o algoritmo SHAP para cada indivíduo, sendo que o efeito da ausência do biomarcador é calculado pela obtenção da distância da previsão original;
- b) se a distância for negativa, o novo valor previsto é maior para a conversão AD do que a previsão original, assim ao indivíduo em questão atribui-se um valor de predição que denota um risco reduzido de conversão em comparação com o risco médio para a população;
- c) calcular uma média de distâncias absolutas entre os indivíduos e gerar uma lista classificada de biomarcadores de acordo com o efeito nas previsões de resultados em todos os indivíduos;
- d) as interações de biomarcadores individuais são calculadas para os analitos de melhor classificação;
- e) o biomarcador em questão (analito de melhor classificação) é então removido do modelo, e cada um dos biomarcadores restantes é removido iterativamente para estimar o efeito de ambos os biomarcadores estarem ausentes ao mesmo tempo;
- f) o efeito da interação é comparado com o efeito aditivo dos biomarcadores individualmente;

g) a diferença entre a soma dos efeitos ao remover os dois biomarcadores individualmente e o efeito da remoção dos dois biomarcadores simultaneamente é calculada para cada indivíduo e a média das distâncias absolutas é usada para encontrar os principais atributos de interação que são utilizados para explicação.

.

## RESUMO

### **“PROCESSO CENTRADO NO HUMANO PARA ELABORAÇÃO DE MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA E USOS”**

A presente tecnologia se refere a um processo centrado no humano para obtenção de modelos baseados em aprendizado de máquina que inclui meios para permitir a um especialista (humano) contribuir para elaboração de modelos, julgando-os. A tecnologia utiliza técnicas como a decomposição do espaço de dados em estruturas locais, do tipo “*backbone*”, formando um conjunto do tipo “*aka backbone features*”; representações vetoriais de explicações de modelos (explicabilidade); agrupamento (“*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 contribuição de especialista humano ao modelo, utilizando metáforas de explicação para julgar características do modelo potencialmente problemáticas, ou até mesmo identificar a ausência de características indispensáveis à modelagem. A tecnologia se aplica 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.