

ALMANAQUE PARA POPULARIZAÇÃO DE CIÊNCIA DA COMPUTAÇÃO

SÉRIE **2** Inteligência
Artificial



VOLUME 6

Aprendizado Baseado em Modelos Probabilísticos



Guilherme de Oliveira Amorim
Albert Santos Barbosa
Maria Augusta Silveira Netto Nunes
Jones Granaty

UNIVERSIDADE FEDERAL DE SERGIPE - UFS

REITOR

Prof. Dr. Angelo Roberto Antonioli

PRO-REITORA

Prof. Dra. Iara Campelo

RESPONSÁVEL PELA PRIMEIRA EDIÇÃO

Albert Santos Barbosa

REVISÃO GERAL

Maria Augusta Silveira Netto Nunes

RESPONSÁVEL PELA SEGUNDA EDIÇÃO (REEDIÇÃO)

Albert Santos Barbosa

Os personagens e as situações desta obra são reais apenas no universo da ficção; não se referem a pessoas e fatos concretos, e não emitem opinião sobre eles.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL
DA UNIVERSIDADE FEDERAL DE SERGIPE

A654a

Aprendizado baseado em modelos probabilísticos [recurso eletrônico] /
Guilherme de Oliveira Amorim... [et al.]. – 2. ed. – Porto Alegre :
SBC ; São Cristóvão : UFS, 2017.

24 p. : il. – (Almanaque para popularização de ciência
da computação. Série 2, Inteligência artificial ; v. 6)

ISBN 978-85-7669-387-1

1. Aprendizado do computador. 2. Inteligência artificial. 3.
Probabilidades. I. Amorim, Guilherme de Oliveira. II. Série.

CDU 004.85(059)



Cidade Universitária José Alóísio de Campos
CEP - 490100-000 - São Cristóvão - SE

Almanaque para popularização de Ciência da Computação

Série 2: INTELIGÊNCIA ARTIFICIAL Volume 6: APRENDIZADO BASEADO EM MODELOS PROBABILÍSTICOS

Sociedade Brasileira de Computação - SBC
Porto Alegre - RS

Autores

Guilherme de Oliveira Amorim
Albert Santos Barbosa
Maria Augusta Silveira Netto Nunes
Jones Granatyr

Realização
Universidade Federal de Sergipe

São Cristóvão – Sergipe
2017

Apresentação

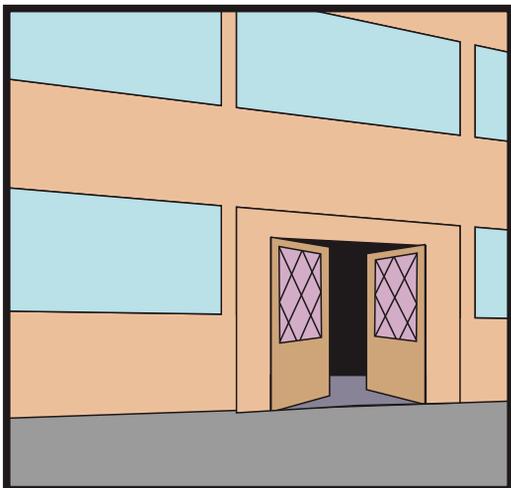
Essa cartilha foi desenvolvida pelo projeto de Bolsa de Produtividade CNPq-DTII n°306576/2016-3, coordenado pela prof^a. Maria Augusta S. N. Nunes em desenvolvimento no Departamento de Computação (DCOMP)/Programa de Pós-graduação em Ciência da Computação (PROCC) – UFS. É também vinculado à projetos de extensão, Iniciação Científica e Tecnológica para popularização de Ciência da Computação em Sergipe apoiado pela PROEX, COPES e CINTTEC/UFS. O público alvo das cartilhas são jovens pré-vestibulandos e graduandos em anos iniciais. O objetivo é fomentar ao público sergipano e nacional o interesse pela área de de Ciência da Computação.

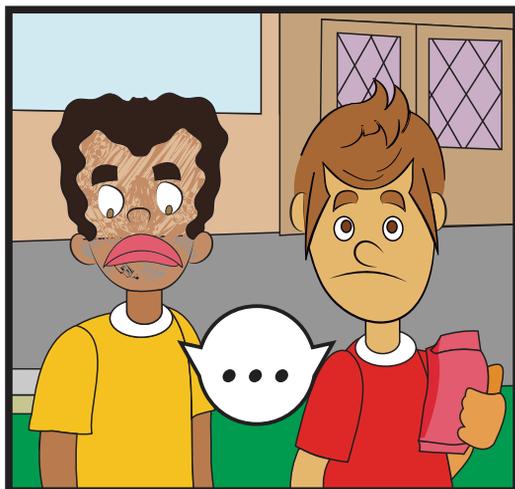
As cartilhas da série sobre Inteligência Artificial descrevem a área da Ciência da Computação que busca simular a inteligência humana por meio de dispositivos e softwares. Essa cartilha mostra uma linha de estudo chamada de Aprendizado Baseado em Modelos Probabilísticos (ABMP). O ABMP é usado quando se quer que o computador aprenda, e um exemplo prático é a classificação de pessoas como crianças, adultas e idosas utilizando fotos.

(Maria Augusta Silveira Netto Nunes)

O toque para o intervalo de aula soa.

OOOCHIIIIIOOCHIIIIII
OOCHIIIIIOOCHIIIIII









Um dos próximos assuntos que vamos abordar em sala de aula é sobre probabilidade.

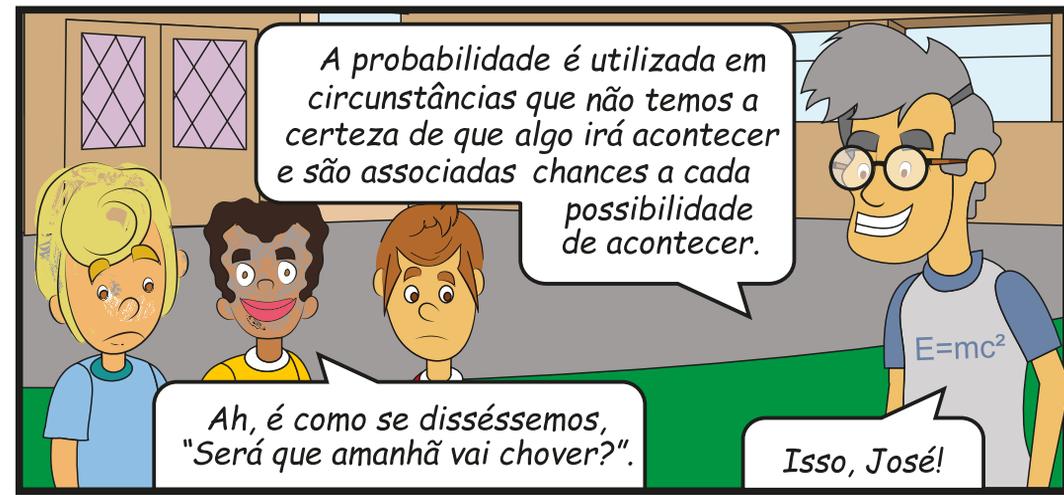
Vocês poderiam me dizer ou dar algum exemplo de que seria probabilidade?



É tentar prever as coisas, professor?



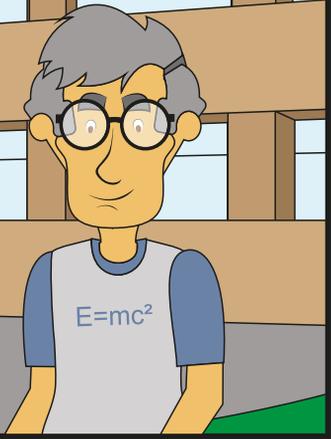
É uma boa resposta, João.



A probabilidade é utilizada em circunstâncias que não temos a certeza de que algo irá acontecer e são associadas chances a cada possibilidade de acontecer.

Ah, é como se disséssemos, "Será que amanhã vai chover?".

Isso, José!



É tentar adivinhar algo que irá acontecer. Ou mesmo, "Qual a probabilidade de José ter tirado uma bala com sabor cereja do saco de balas?".



Ah, entendi! Seria algo como qual a probabilidade de sair balas de cereja ou limão? Não é isso?



Exatamente José! Com isso temos o conjunto de nossas probabilidades, que nada mais é do que provar uma afirmação que fazemos sobre algo que pode acontecer.



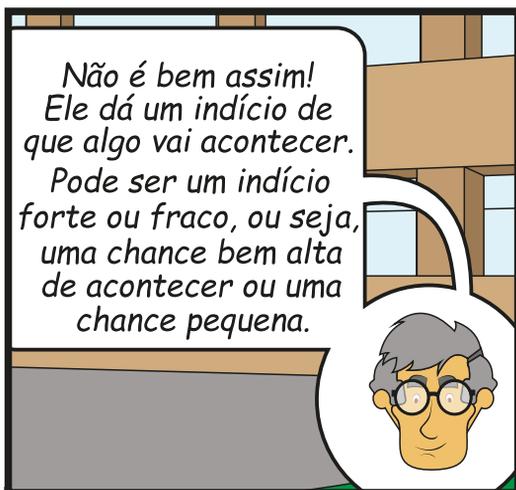
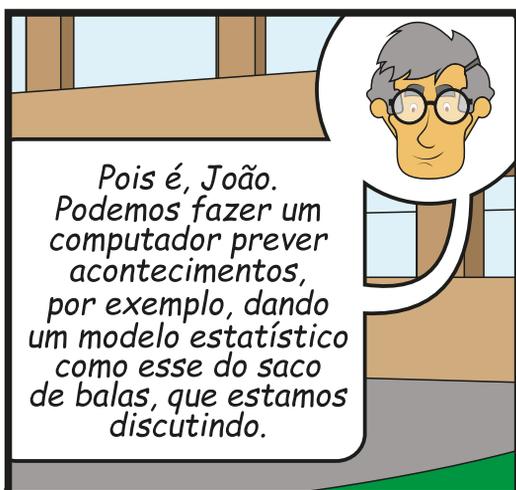
Como assim?

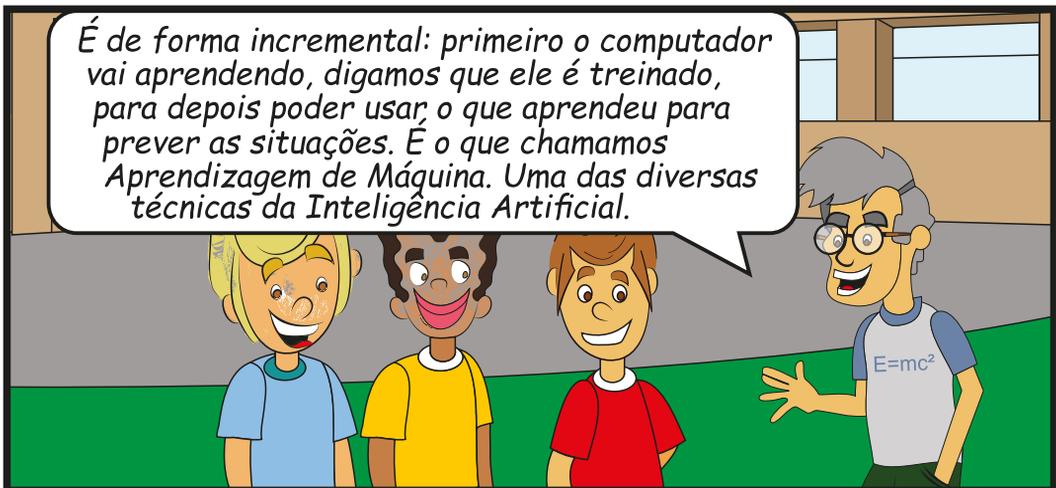


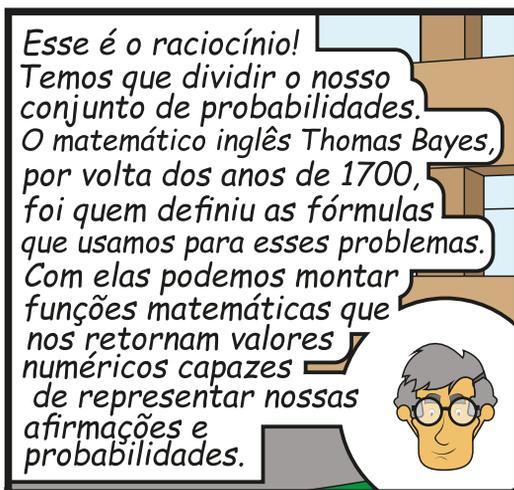
Bom, vamos supor que todas as balas que serão retiradas do saco de balas sejam de limão.

Essa é nossa afirmação. Agora precisamos prová-la, a partir de cada bala que retiramos do saco.

Se todas as balas retiradas forem de limão, conseguimos comprovar nossa afirmação.

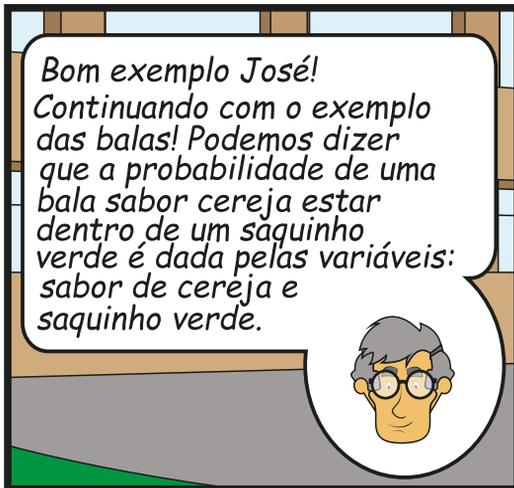




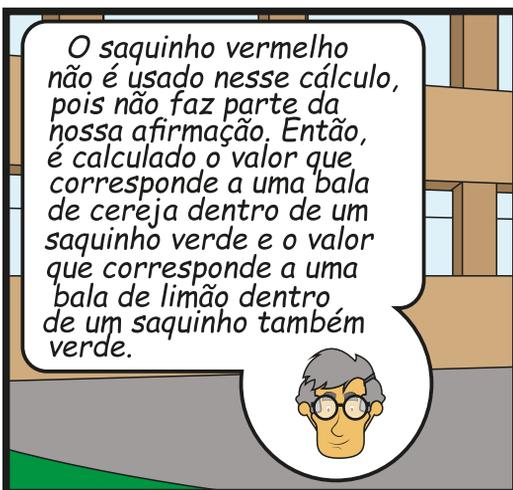




O jogo de cara e coroa funciona assim, não é? Temos cinquenta por cento de chances do lado da moeda ser o que escolhemos, a cada jogada.



Bom exemplo José!
Continuando com o exemplo das balas! Podemos dizer que a probabilidade de uma bala sabor cereja estar dentro de um saquinho verde é dada pelas variáveis: sabor de cereja e saquinho verde.



O saquinho vermelho não é usado nesse cálculo, pois não faz parte da nossa afirmação. Então, é calculado o valor que corresponde a uma bala de cereja dentro de um saquinho verde e o valor que corresponde a uma bala de limão dentro de um saquinho também verde.



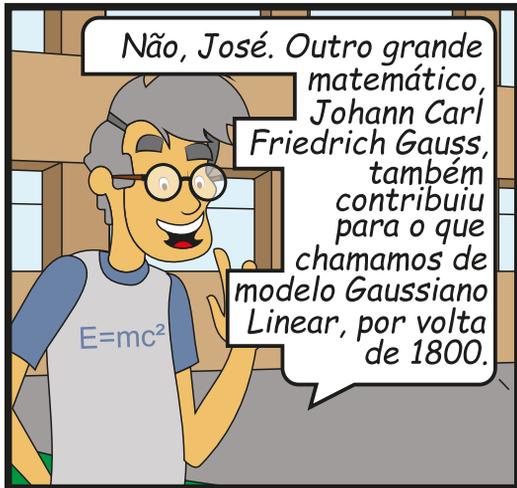
Entendi! Assim, temos os valores para cada variável que queremos analisar, e com isso observamos qual a hipótese com maior probabilidade de acontecer.



Perfeito, João!



Mas, professor, só existe esse método? Só Bayes que ajudou os computadores a aprenderem?



Não, José. Outro grande matemático, Johann Carl Friedrich Gauss, também contribuiu para o que chamamos de modelo Gaussiano Linear, por volta de 1800.

O modelo proposto por Gauss trabalha com média e desvio padrão. A média é algo com que nos depáramos no dia-a-dia, como quando queremos saber a nossa média do bimestre. A média de um conjunto de valores numéricos é a soma de todos estes valores dividida pelo número de elementos somados.

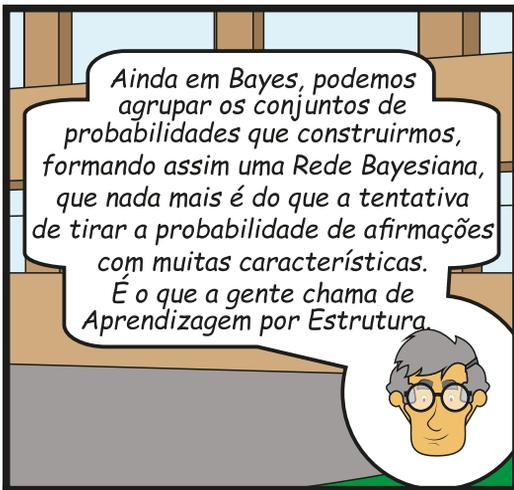
Já o desvio padrão, é obtido através da média e nos diz se há muita variação de cada valor em relação à média.



Através dos dois representamos a maior probabilidade para uma afirmação, ou seja, a maior chance de um evento, como chover amanhã, acontecer.

Com apenas duas características? Só isso? E porque Bayes utilizava vários?

Isso, José. Com apenas dois! Bayes se utilizou de vários, no nosso exemplo das balas até dois, pois ele tratou diretamente com os eventos, ou seja, a hipótese de uma bala de cereja estar em um saquinho verde, por exemplo. Já Gauss utiliza uma amostra do nosso conjunto.





HAHAHA!

Existem variáveis desconhecidas, e o desafio, aqui, é determinar as variáveis ocultas a partir das variáveis que já temos e, as novas variáveis encontradas podem então serem usadas para realizar novas análises.



Como assim, professor?



São técnicas mais avançadas. E preciso ter um conhecimento maior sobre a Ciência da Computação e, principalmente, sobre Inteligência Artificial.



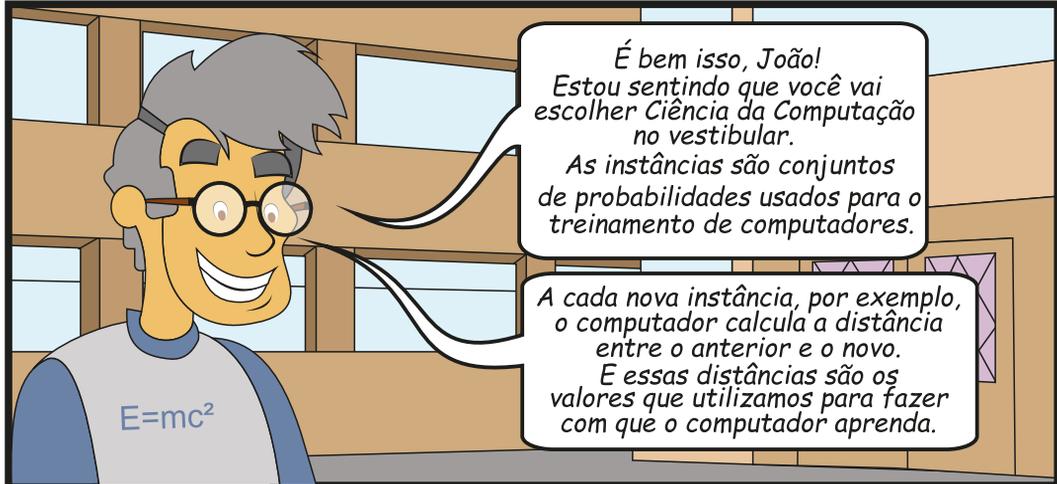
Mas vamos lá: podemos aplicar um sistema de computação que forma agrupamentos com nossos dados e tenta distinguir categorias dentro de uma coleção de dados.

Então, com a entrada de um dado novo, o computador aprende a classificá-lo mais facilmente.



Encontrar grupos de objetos tal que objetos em um grupo são similares ou relacionados uns aos outros e diferentes ou não relacionados a objetos em outros grupos, por exemplo, separar documentos de texto por nome, data, tipo de documento, ou mesmo, separar um conjunto de pessoas em torcedores de um time ou de outro.





É bem isso, João!
Estou sentindo que você vai
escolher Ciência da Computação
no vestibular.
As instâncias são conjuntos
de probabilidades usados para o
treinamento de computadores.

A cada nova instância, por exemplo,
o computador calcula a distância
entre o anterior e o novo.
E essas distâncias são os
valores que utilizamos para fazer
com que o computador aprenda.



Entenderam?
Quanto menor a distância
entre um modelo e outro,
temos mais certeza de
que determinado dado é
de um grupo ou de outro.

Isso acontece
com os filmes que
a Netflix nos
recomenda, por exemplo.

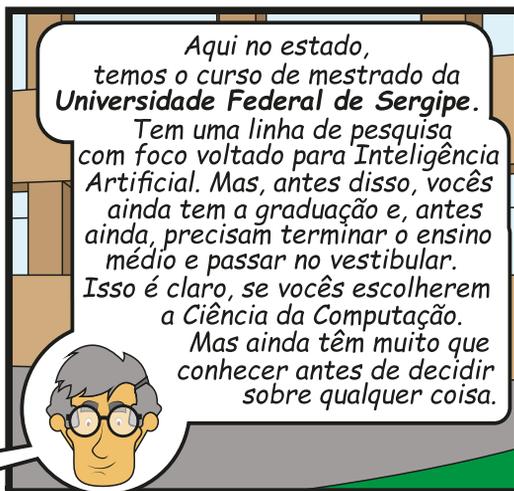


Nossa! É muito extensa
a área de Aprendizado
de Máquina, não é?



Onde podemos estudar
sobre essas coisas?

É uma área vasta, sim.
Existem diversas técnicas e
muitas ainda estão surgindo.



Aqui no estado,
temos o curso de mestrado da
Universidade Federal de Sergipe.
Tem uma linha de pesquisa
com foco voltado para Inteligência
Artificial. Mas, antes disso, vocês
ainda tem a graduação e, antes
ainda, precisam terminar o ensino
médio e passar no vestibular.
Isso é claro, se vocês escolherem
a Ciência da Computação.
Mas ainda têm muito que
conhecer antes de decidir
sobre qualquer coisa.



Fim!

Bibliografia

RUSSELL, Stuart; NORVIG, Peter; INTELLIGENCE, Artificial. A modern approach. Artificial Intelligence. Prentice-Hall, Englewood Cliffs, v. 25, 1995.

Mais cartilhas em:

<http://almanaqsdacomputacao.com.br/index.html>

<http://meninasnacomputacao.com.br/gutanunes/publication.html>

<http://meninasnacomputacao.com.br/>

Sobre os autores

GUILHERME DE OLIVEIRA AMORIM

Bolsista FAPITEC mestrando em Computação Inteligente com ênfase em Análise de Sentimento e Personalidade pela Universidade Federal de Sergipe. Atuou na área de Análise de Testes, como Tester pelo Cin, Centro de Informática da Universidade Federal de Pernambuco, em parceria com a Motorola. Atuou, também, como desenvolvedor de software pela Universidade Federal de Alagoas, Campus Arapiraca, em projetos com ênfase em Biotecnologia e Desenvolvimento de Aplicações Desktop, ambos utilizando a Plataforma Java e sistema gerenciador de banco de dados PostgreSQL.

ALBERT SANTOS BARBOSA

Bolsista do Projeto Popularização da Ciência da Computação em Sergipe PAEX.

Bolsista PAEX, Criando ilustrações das cartilhas informativas para popularização da Ciência da Computação em Sergipe. Graduando em Design Gráfico pela Universidade Federal de Sergipe (2012.1). Possui experiência na área do Design gráfico, com ênfase em ilustração.

MARIA AUGUSTA SILVEIRA NETTO NUNES

Bolsista de Produtividade Desen. Tec. e Extensão Inovadora do CNPq - Nível 2 - CA 96 - Programa de Desenvolvimento Tecnológico e Industrial

Professor Adjunto IV do Departamento de Computação da Universidade Federal de Sergipe. Membro do Programa de Pós-graduação em Ciência da Computação (PROCC) na UFS. Pós-doutora em Propriedade Intelectual no Instituto Nacional de Propriedade Industrial (INPI). Doutora em "Informatique pela Université de Montpellier II - LIRMM em Montpellier, França (2008). Realizou estágio doutoral (doc-sanduíche) no INESC-ID-IST Lisboa- Portugal (ago 2007-fev 2008). É mestre em Ciência da Computação pela Universidade Federal do Rio Grande do Sul (1998) e possui graduação em Ciência da Computação pela Universidade de Passo Fundo (1995). Possui experiência acadêmico-tecnológica na área de Ciência da Computação e Inovação Tecnológica/Propriedade Intelectual. Atualmente, suas pesquisas estão voltadas, principalmente na área de inovação tecnológica usando Computação Afetiva na tomada de decisão Computacional. Atua também em Inovação Tecnológica, Propriedade Intelectual capacitando empresários na área de TI e fornecendo consultoria em Registro de Software e patente.

Lattes: <http://lattes.cnpq.br/9923270028346687>

JONES GRANATYR

Doutorando em Informática bolsista da CAPES e Mestre em Informática bolsista do CNPq (2011), ambos na área de Inteligência Artificial. No mestrado trabalhou dentro do contexto do projeto PAI-L (Piloto Automático Inteligente para Locomotivas) financiado pela FINEP (Financiadora de Estudos e Projetos), o qual teve como objetivo o desenvolvimento de um software inteligente para condução de locomotivas de carga. Possui Especialização em Segurança em Redes de Computadores e Banco de Dados Distribuídos (2007) e Graduação em Sistemas de Informação (2004). Trabalha em projetos de pesquisa relacionados a área Inteligência Artificial, tais como Sistemas Especialistas, Mineração de Dados, Mineração de Textos, Sistemas Multiagente, Aprendizagem de Máquina e Computação Afetiva. Entre os anos de 2011 e 2012 foi bolsista de produtividade nível B1 do CNPq, trabalhando com Raciocínio Baseado em Casos no software JUSTINIA (Justiça Inteligente Apoiada em Inteligência Artificial). Em 2010 teve sua empresa contemplada com financiamento modalidade subvenção econômica da FINEP vinculado ao projeto PRIME (Primeira Empresa Inovadora). Atualmente é líder do grupo de pesquisa de Inteligência Computacional do CNPq e Universidade do Contestado. É também membro do grupo de pesquisa de Agentes de Software da Pontifícia Universidade Católica do Paraná. Leciona na Universidade do Contestado - SC no curso de Ciência da Computação e na União de Ensino do Sudoeste do Paraná - UNISEP de Francisco Beltrão - PR no curso de Sistemas de Informação. ação de pessoas como crianças, adultas e idosas utilizando fotos.

Agradecimentos

Ao CNPq, CAPES, SBC, BICEN, DCOMP, PROCC, PROEX e NIT/UFS.



PROCC
Programa de Pós-Graduação em
Ciência da Computação/UFS



PROEX



ISBN 978-857669387-1



9

788576

693871