Fronteiras da Engenharia de Software - Podcast

Episódio 6, Temporada 2: Aprendizado de Máquina em Engenharia de Software, com Maurício Aniche (TU Delft)

Host: Adolfo Neto (UTFPR)

Co-host: Ingrid Nunes (UFRGS)

Equipe: Danilo Ribeiro (Zup), Leonardo Fernandes (IFAL), Fabio Petrillo (Univ. Quebec), Gustavo Pinto (UFPA)

# Script do Episódio

## Parte 1: Apresentação [5 min, estimativa]

[ADOLFO] Hoje no Fronteiras da Engenharia de Software vamos conversar com Mauricio Aniche, professor na Universidade Tecnológica de Delft.

[ADOLFO] Tudo bem, Maurício? Você pode se apresentar para as pessoas que nos escutam?

CORTAR RESPOSTA 1

## Parte 2 - Tema do Episódio

[INGRID] O tema deste episódio é aprendizado de máquina em Engenharia de Software. Antes de a gente dar o foco em Engenharia de Software, é bacana que nossos ouvintes saibam o que é aprendizado de máquina, principais conceitos, e como isso se difere de inteligência artificial e ciência de dados. Você pode dar um panorama geral sobre o assunto?

[ADOLFO] Agora, focando em Aprendizado de Máquina \*em Engenharia de Software\*. A gente tem ouvido esse termo em diversas conferências. Isso é uma área de pesquisa? É um tipo de técnica? Como isso é definido?

CORTE DEPOIS DA RESPOSTA

[INGRID] Então, aprendizado de máquina para Engenharia de Software não é um problema a ser resolvido, como tradicionalmente temos em pesquisa na Engenharia de Software. Isso é válido? O caminho não é identificar problemas e, no caso de machine learning ser uma solução adequada, utilizar isso?

[ADOLFO] Antes de machine learning para a Engenharia de Software, os sistemas de recomendação da Engenharia de Software já tinham ganhado popularidade. São duas coisas com nomes diferentes? Qual a relação entre esses dois tópicos?

[INGRID] Considerando então que a ideia então é explorar o uso de aprendizado de máquina para resolver problemas de Engenharia de Software. Você poderia dar exemplos de classes de problemas que são endereçados usando aprendizado de máquina nessa área? Esses problemas estão relacionados com as diversas atividades do desenvolvimento de software, isto é, especificação, projeto e implementação, validação e verificação e evolução?

[ADOLFO] E falando agora de técnicas de aprendizado de máquinas que são utilizadas, quais são as mais utilizadas? Por exemplo, a gente ouve falar do uso de redes neurais e outras técnicas bastante populares, não só em Engenharia de Software.

[INGRID] Dentro do contexto de Engenharia de Software, nós temos também a área de Engenharia de Software baseada em Busca. Você inclusive tem trabalho nesta linha. Até onde sei, SBSE se baseia em algoritmos genéticos, que são mais úteis quando a gente não tem um norte da onde chegar. A gente tem componentes aleatórios visando conseguir convergir para um resultado. Faz sentido usar isso na Engenharia de Software? Digo isso porque na Engenharia de Software nós sabemos onde queremos chegar de forma geral, como por exemplo, ter software mais modular, casos de testes que têm maior probabilidade de encontrar erros, etc. Antes de responder, poderias por favor dar uma breve explicação do que são os algoritmos genéticos?

[ADOLFO] Pensando agora no uso dessas abordagens de aprendizado de máquina em engenharia de software em projetos de software. Simplesmente fazer modelos capazes de predizer alguma coisa não tem utilidade direta. Como transformar essas contribuições em ferramentas? Que outras questões devem ser pensadas, como usabilidade de ferramentas?

[INGRID] Agora fazendo uma pergunta do tipo advogado do diabo. Existe um artigo do Dietmar Jannach e Malte Ludewig, no qual os autores mostraram que várias técnicas que são o estado da arte e que usam redes neurais têm resultados piores do que KNN, se o algoritmo for parametrizado adequadamente. Você acha que existe uma pressão para usar soluções rebuscadas para problemas quando soluções mais simples poderiam ser empregadas?

## Parte 3: Pesquisa

[ADOLFO] Você tem trabalhos recentes com estudos de caso na Adyen, uma empresa holandesa de pagamentos. Você pode contar como foram estes estudos de caso e qual foi o aprendizado?

[INGRID] Um outro contexto em que você usou aprendizado de máquina é para identificar locais no código onde logs devem ser colocados. Você poderia falar mais sobre sobre este trabalho?

CORTE

Sistemas de ML são complexos para colocar em produção. Alguns chamam essas práticas de MLOps. Você teria algumas casos para discutir sobre como colocar esses modelos em produção?

[ADOLFO] A gente falou até agora sobre como aprendizado de máquina pode contribuir para a área de engenharia de software. E o contrário? Desenvolver partes de sistemas de software que envolvem aprendizado de máquina é difícil? Como então a engenharia de software pode contribuir para a área de aprendizado de máquina? Você inclusive tem um trabalho sobre isso.

## Parte 4: Outras perguntas

[INGRID] Tem um outro trabalho seu, não no contexto de aprendizado de máquina, que eu curto muito. O trabalho investiga code smells em arquiteturas model-view-controller (MVC). Eu já tive quase que um "infarto" quando vi blogs por aí de gente argumentando que o model são os dados e o controller tem a lógica de negócio. Primeiro, não é incomum em projetos de software ver classes de dados porque elas são persistidas no banco de dados e em algum momento as pessoas acharam que não podem colocar nada nessas classes além de getters e setters. Segundo, é o perigo de informações erradas serem difundidas e se tornarem "verdades". Qual a tua visão sobre esses dois assuntos?

CORTE

[ADOLFO] Na instituição em que você trabalha, você ensina para turma de um número muito grande de alunos. Isso traz um grande desafio muito grande na hora de corrigir trabalhos e avaliações. Você pode compartilhar a sua experiência nisso? Você inclusive tem um artigo onde você reportou o caso dos alunos se auto-avaliarem, correto?

## Parte 5: Próxima Fronteira da ES [3 min, estimativa]

[ADOLFO] Para você, qual é a próxima fronteira da engenharia de software? (pode ser algo que você acha que vai acontecer ou que você gostaria que acontecesse em nossa área)

## Parte 6: Encerramento

Adolfo agradece e passa para o(a) entrevistado(a).

Ingrid fecha o episódio.

# Texto para divulgação

Conversamos com Maurício Aniche, pesquisador na TU Delft.

Sites de Maurício

* <https://www.mauricioaniche.com/>

Links Citados

| Aprendizado de Máquina em Engenharia de Software, com Maurício Aniche (TU Delft) |
| --- |
| Neste episódio, conversamos com Maurício F. Aniche, professor na Universidade Tecnológica de Delft (TU Delft). O tema foi Aprendizado de Máquina em Engenharia de Software.  A página do Maurício é <https://www.mauricioaniche.com/>. Ele está no Twitter como [@mauricioaniche](https://twitter.com/mauricioaniche).  Artigos mencionados:   * Jeroen Castelein, Maurício Finavaro Aniche, Mozhan Soltani, Annibale Panichella, Arie van Deursen: Search-based test data generation for SQL queries. ICSE 2018. * Ludewig, M., Mauro, N., Latifi, S. and Jannach, D.: "Performance Comparison of Neural and Non-Neural Approaches to Session-based Recommendation". In: Proceedings of the 2019 ACM Conference on Recommender Systems (RecSys 2019). Copenhagen, 2019 * Henk Grent, Aleksei Akimov, Maurício Aniche: Automatically Identifying Parameter Constraints in Complex Web APIs: A Case Study at Adyen. ICSE (SEIP) 2021. * Casper Schröder, Adriaan van der Feltz, Annibale Panichella, Maurício Aniche: Search-Based Software Re-Modularization: A Case Study at Adyen. ICSE (SEIP) 2021. * Jeanderson Cândido, Jan Haesen, Maurício Aniche, Arie van Deursen: * An Exploratory Study of Log Placement Recommendation in an Enterprise System. MSR 2021: 143-154 * Bart van Oort, Luis Cruz, Maurício Aniche, Arie van Deursen: * The Prevalence of Code Smells in Machine Learning projects. WAIN@ICSE 2021 * Mauricio Finavaro Aniche, Gabriele Bavota, Christoph Treude, Marco Aurélio Gerosa, Arie van Deursen: Code smells for Model-View-Controller architectures. Empir. Softw. Eng. 23(4): 2121-2157 (2018) * Maurício Aniche, Frank Mulder, Felienne Hermans: Grading 600+ Students: A Case Study on Peer and Self Grading. ICSE (SEET) 2021: 211-220   Nossa equipe é formada por:   * Leonardo Fernandes (IFAL) * Gustavo Pinto (UFPA) * Fabio Petrillo (UQAC) * Danilo Monteiro (Zup Innovation) * Adolfo Gustavo Serra Seca Neto (UTFPR) - co-host deste episódio * Ingrid Nunes (UFRGS) - co-host deste episódio   Nosso site é<https://fronteirases.github.io/>.  A edição deste episódio foi feita pela Zup Innovation. |