

II WORKSHOP EM DATA SCIENCE - UFLA

08/01 – Segunda-feira

	PALESTRA	PALESTRANTE
14:00	Detecting and verifying online disinformation: how NLP and data analysis can help?	Carolina Scarton
16:30	---	Ana Carolina
20:15	Ciência de Dados e PLN: Desafios e Oportunidades na Extração da Informação.	Daniela Barreiro Claro
19:00	Como "desenhar" uma solução de machine learning?	Rafael Tuma Guariento

09/01 – Terça-feira

	PALESTRA	PALESTRANTE
14:00	NLU: como fazer as máquinas "entenderem" linguagem humana	Fernando Antônio Asevedo Nobrega
15:15	---	Jonathan Cabreira
16:30	---	Eric Fernandes de Mello Araújo
19:00	Stop the Clock: Are Timeout Effects Real?	Renato Martins Assunção
20:15	Mercado de Dados na Mobilidade como um Serviço: principais tendências	Arthur de Miranda Neto



RESUMOS

- 1) Detecting and verifying online disinformation: how NLP and data analysis can help? **Palestrante:** Carolina Scarton

Resumo: Hoje em dia, desinformação é algo que encontramos recorrentemente em websites e medias sociais. Com o avanço da pandemia do COVID-19, um fenômeno nomeado “infodemic” também ocorreu: desinformação está se propagando tão rapidamente quanto o vírus. Desde o começo da pandemia, jornalistas tem dedicado grande parte de seus esforços na tarefa de checagem de conteúdo. Porém, a quantidade de conteúdo que necessita verificação é muito maior do que a quantidade de pessoas fazendo a verificação. Neste cenário, técnicas de Processamento de Língua Natural (PLN) e outras áreas no escopo da Ciência de Dados tem grande potencial de aplicação. Nesta palestra, vou apresentar nosso entendimento sobre desinformação no âmbito online, com ênfase nos desafios para o desenvolvimento de ferramentas computacionais. O projeto europeu WeVerify (<https://weverify.eu>) também será apresentado, com foco no trabalho que utiliza PLN desenvolvido na Universidade de Sheffield (UK).

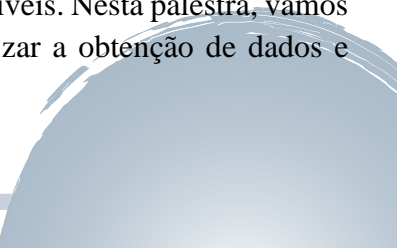
- 2) Ana Carolina


- 3) Como "desenhar" uma solução de machine learning? **Palestrante:** Rafael Tuma Guariento

Resumo: Quando aprendemos uma nova ferramenta, temos a tendência de "encaixar" essa ferramenta em tudo que vemos. Mas o mundo real não vem pra gente falando "olha, eu preciso de um modelo de classificação, sem necessidade de estimação da probabilidade, com essa variável resposta, que possui um balanceamento de 30% entre as classes". O mundo real fala "eu tenho essa DOR. Me ajuda a resolver?" Como vamos usar nosso portfólio de conhecimentos e ferramentas pra resolver, cabe a nós decidirmos, e deve, muito mais importante do que algo complexo com as últimas técnicas, ser útil. Nessa palestra eu falo sobre a técnica que desenvolvi ao longo de anos como cientista de dados sobre como transformar um problema de negócio em um desenho de solução, que pode ou não contemplar modelos, eventualmente mais de um modelo, e como transformar as saídas em algo útil e visualizável pelo seu cliente.

- 4) Ciência de Dados e PLN: Desafios e Oportunidades na Extração da Informação. **Palestrante:** Daniela Barreiro Claro

Resumo: A intersecção da Ciência de Dados e do Processamento de Linguagem Natural potencializa diversas novas abordagens para analisar dados, principalmente considerando os dados não-estruturados. A grande quantidade desses dados na Web demanda novos mecanismos de extração de informação automatizados, que possam auxiliar diversas aplicações. Dentro desse cenário, os tratamentos monolíngues e multilíngues têm sido amplamente analisados a fim de usufruir da vasta quantidade de textos disponíveis. Nesta palestra, vamos abordar algumas das nossas soluções de Extração de Informação para automatizar a obtenção de dados e apresentar alguns dos desafios e aplicações dessa área.



- 
- 5) NLU: como fazer as máquinas "entenderem" linguagem humana. **Palestrante:** Fernando Antônio Asevedo Nobrega

Resumo:

- 6) **Palestrante:** Jonathan Cabreira

Resumo:

- 7) **Palestrante:** Eric Fernandes de Mello Araújo

Resumo:

- 8) Stop the Clock: Are Timeout Effects Real? **Palestrante:** Renato Martins Assunção

Resumo: Timeout é uma curta interrupção durante os jogos esportivos, usada para comunicar uma mudança na estratégia, para dar aos jogadores um descanso ou para parar um fluxo negativo no jogo. Seja qual for o motivo, os treinadores esperam uma melhora no desempenho de sua equipe após um tempo limite. Mas quão eficazes são esses tempos limite para fazer isso? A média simples das diferenças entre as pontuações antes e depois dos tempos limite foi usada como evidência de que existe um efeito e que é substancial. Neste artigo, mostramos esta análise é incorreta e que, de fato e com surpresa, o efeito do timeout é nulo. Usamos um modelo gráfico causal e uma grande base de dados de jogos de basquete da National Basketball Association (NBA) para comprovar nossas afirmações. Este trabalho é uma co-autoria com Niander Neves (Microsoft) e Pedro Vaz de Melo (DCC-UFMG).

- 9) Mercado de Dados na Mobilidade como um Serviço: principais tendências. **Palestrante:** Arthur de Miranda Neto

Resumo: A mobilidade é uma necessidade humana assim como comer e dormir. Ela é parte integrante de um sistema logístico eficiente. No centro das transações desses mercados estão os dados.

