



















Fatec
Ourinhos











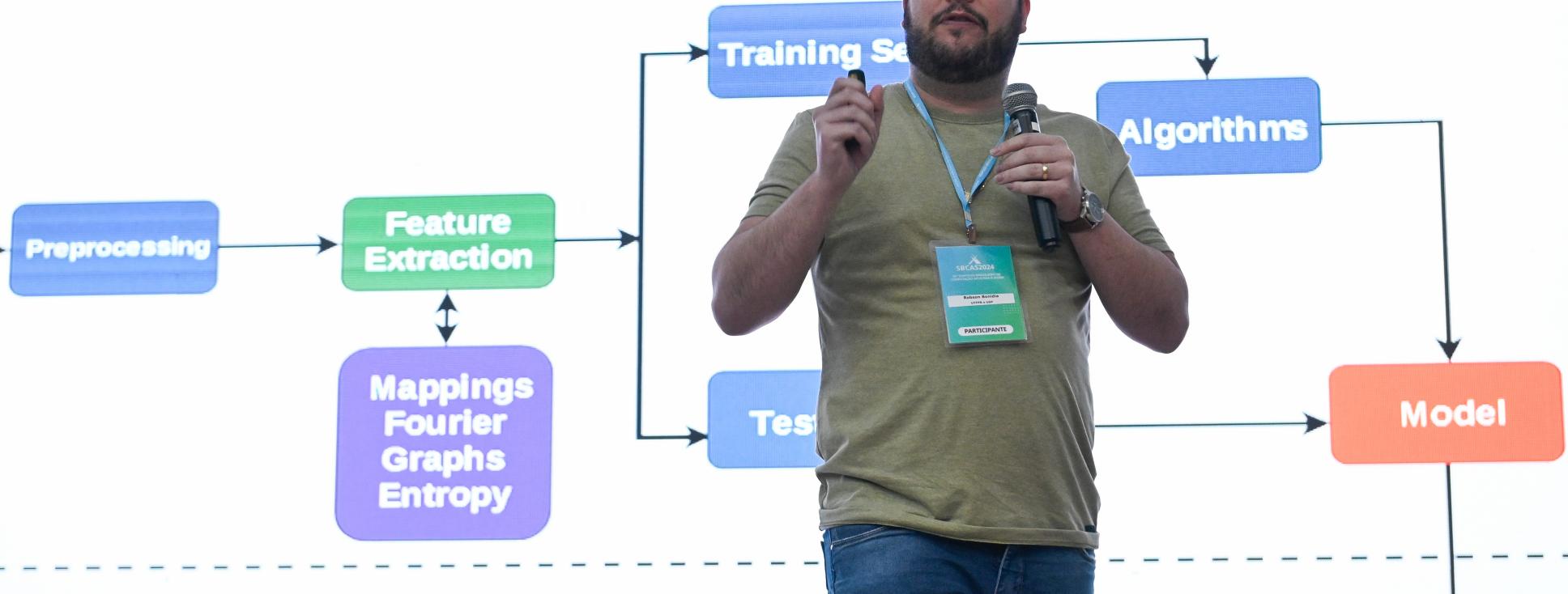
RESERVADO

Democratization

ing

I s

ting





25

20



virou Notícia! Fatec Ourinhos

Fatec Ourinhos conquista o 1º Lugar no Falling Walls Lab Brazil de 2023

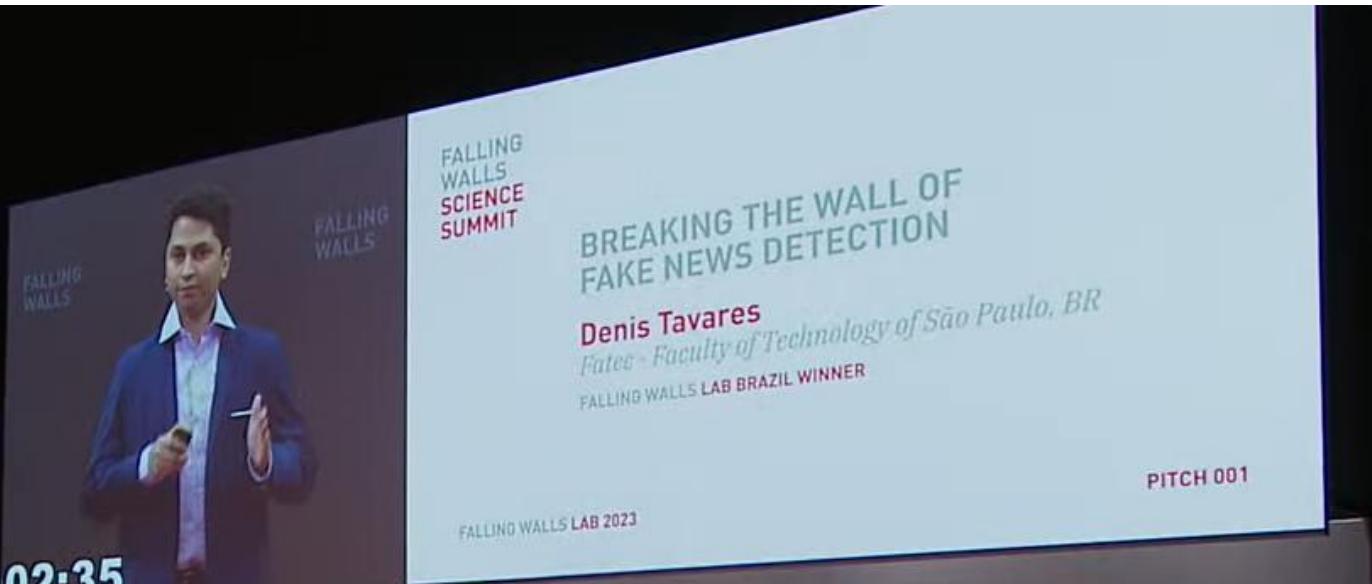




GOVERNO DO
SÃO P







GOVERNO DO ESTADO DE
SAO PAULO

Centro Paula Souza



25 anos



PODER LEGISLATIVO





BioPrediction: Democratizando o Aprendizado de Máquina no Estudo de Interações Moleculares

Florentino^a, Bruno, Sanches^b, Natan, Bonidia^b, Robson, Carvalho^b, André
 (a) Instituto de Física de São Carlos (IFSC-USP), (b) Instituto de Ciências Matemáticas e de Computação (ICMC-USP)

CONTEXTO

o crescente número de sequências biológicas armazenadas em bancos de dados, existe uma grande fonte de informações que pode beneficiar os setores, uma vez que esses dados podem ser usados para inferir características de outros problemas similares, mas que ainda estão em aberto. Nesse contexto, algoritmos de Aprendizado de Máquina (AM) podem decifrar padrões nesses dados e portanto extrair novas informações úteis para pesquisadores de áreas relacionadas à biologia molecular. Entretanto, é necessário algum conhecimento técnico em AM para construir modelos capazes de fazer previsões confiáveis dado a natureza não estruturada dos dados.

OBJETIVOS

framework Fim-a-fim baseado em Aprendizado de Máquina automatizado, chamado BioPrediction (uma extensão da BioAutoML [1]), capaz de identificar interações implícitas entre moléculas biológicas, por exemplo, pares de RNA longo (lncRNAs) e proteínas, sem a necessidade de conhecimento prévio ou de ponta a ponta. Adicionalmente, pretende-se fornecer uma saída útil aos usuários por meio de diversas formas de visualização das métricas do modelo, da topologia das interações e da sua interpretabilidade.

Figura 2: Comparação da métrica AUC nos diferentes datasets.

AUC dos datasets por ferramenta analisada

No módulo de interpretabilidade temos gráficos que visam esclarecer alguns padrões na tomada de decisão do modelo. Temos uma noção global das previsões na figura 3, nesse caso vemos que as características mais importantes são de proteínas e não de RNAs. Ademais, pesquisadores interessados podem procurar por características com alta separação entre valores altos e baixos (pontos vermelhos e azuis) e investigar melhor esses padrões.

beeswarm_graph

Figura 3: Gráfico da biblioteca SHAP que relaciona valor assumido pela característica com a contribuição na previsão da classe final.

CONCLUSÕES

Por fim, o BioPrediction representa um possível avanço no campo da predição de interações, ao democratizar o acesso ao tratamento de dados biológicos e treinamento automatizado de modelos de AM, de forma que esses possam desempenhar competitivo quando comparados aos criados por especialistas.

REFERÊNCIAS

[1] Bonidia, et al. BioAutoML: automated feature engineering and metalearning to predict noncoding RNAs in bacteria. *Briefings in Bioinformatics*, 23(4), 2022.

APOIADORES

Arcos | STARTUPS

