Reunião 11/08

Participantes - João, William, Giovanni, Fabricio, Thiago (iFood)

Descrição do problema - A empresa possui modelos de machine learning (já em produção) que apresentam comportamentos inesperados em face de dados do mundo real, especialmente em situações adversariais, como fraudadores tentando burlar esses modelos. Um exemplo seria de um modelo de classificação de fraude cujo comportamento de uma das features deveria ser linear, ou seja, quanto maior o valor dessa feature, maior a chance de fraude, e vice-versa, porém a partir de um certo valor esse comportamento se altera e o modelo começa a prever de forma equivocada.

Proposta – Um programa (quais estratégias serão utilizadas ainda não foram definidas) que identifique esses bugs, tendo como entrada as features utilizadas pelo modelo e o modelo em si (caixa preta).

Primeiros passos – Buscar literatura que trate de situações semelhantes (em especial para modelos de classificação com dados estruturados/tabulares), estudar o funcionamento de testes de software (unitário, integração) e escolher um dataset que será utilizado no projeto.