

Ética, Data Science e a GDPR

Gabriel Barbosa Gaspar Veloso
João Vasconcelos de Souza Neto





Titulo desse slide

- Topico 1, com textos enxutos
 - Subtopicos
 - Sub
 - ...
- Topico 2 se relacionado ao titulo do slide
- ...



A promessa da Inteligência Artificial

- People Analytics
 - Data Science na hora de recrutar ou analisar funcionários
- Classificação de Pessoas
 - Problemas clássicos como concessão de crédito
- Predição de de Comportamentos
 - Prever o comportamento da bolsa de valores, do preço de um produto no futuro, etc



O que é ser ético em Data Science?

- De onde vem os dados?
 - Foram coletados apropriadamente?
 - Refletem a realidade?
- A quem influenciam?
 - Como influenciam essas pessoas?
- Para que fins estão sendo utilizados?
 - Os impactos sociais são positivos?



Impactos negativos de algoritmos errôneos tem se destacado:

- [Coded Gaze](#): visão computacional e tons escuros de pele.
- [Facebook e Cambridge Analytica](#): vazamento de dados pessoais de milhões de usuários.
- Quão grande é o impacto de algoritmos enviesados?



General Data Protection Regulation: Regulamentação para trabalho com dados.

Regulamentação que vigorou em 2018, traz regras para o processamento de dados:

- Sujeitos a coletas devem ser informados.
- Caso solicitem, têm direito a receber explicações sobre decisões automatizadas.
- Evitar vieses e discriminação em conjuntos de dados.
- Minimizar uso de Personally Identifiable Information (PII) e mantê-las seguras.
- Reportar vazamentos de dados em no máximo 72 horas.
- Invalidação de certos conjuntos de dados anteriores à GDPR.
- Objetivos de processamento devem ser claros e não podem ser ultrapassados.



Responsabilidades recaem sobre Cientistas dos Dados.

- Coleta
 - Garantir que trará dados que reflitam a realidade
- Processamento
 - Processamento adequado e transparente
- Manutenção
 - Perpetuar corretude dos dados e modelos



Conjuntos de dados enviesados

- .Dados enviesados levam a discriminação
 - Utilizar AED para garantir que não é o caso
 - Vieses indiretos também são risco
- Linhagem dos dados
 - Verificar modificações sucessivas de dados



Ambiguidade da GDPR

Por ser recente e não se aprofundar grandemente em certas restrições postas, principalmente o “direito a explicação” para consumidores, há controvérsias sobre as verdadeiras implicações da GDPR.

Cuidados devem ser tomados de qualquer forma, mas não existem certezas sobre quão intensamente restritivas serão as imposições da GDPR.

Se levada a um extremo, técnicas como *Deep Learning* podem tornar-se ilegais, a outro, e consumidores podem receber explicações extremamente superficiais sobre os algoritmos de decisão.



Ambiguidade da GDPR

- A que profundidade devem ir explicações?
 - Descrição superficial?
 - Detalhes do treinamento dos algoritmos?
- Incerteza tida por empresas, empregados e juristas.



Ambiguidade da GDPR

- Algoritmos ilegais?
 - Deep Learning pode somente ser descrito.
 - Decision Tree pode ser mais adequado.
- Segredos de negócio devem ser revelados?
 - Ocultar detalhes pode ser de interesse.



Debate e participação ativa dos Cientistas dos Dados.

- Incertezas devem diminuir com novos hábitos
 - Cientistas dos Dados e empresas estabelecerão padrões
- Debate deliberado para gerar abertura
 - Interesse principal deve ser dos consumidores
 - Como inseri-los nas discussões?



Ética e GDPR

- Um passo adiante?
 - Ambiguidade afasta rigidez necessária.
 - Ainda assim, novo caminho.
- Como ativamente promover a ética?
 - Cada um terá papel na formação do novo ambiente.