



Utilizando NLP para prever detratores de pesquisas de NPS

Palestrante: Milton Ossamu

Patrocinadores Diamante

FORNETIX



SecuPi



Qexpert



dataEX



**Business
Insights &
Governance**

Apoio Institucional



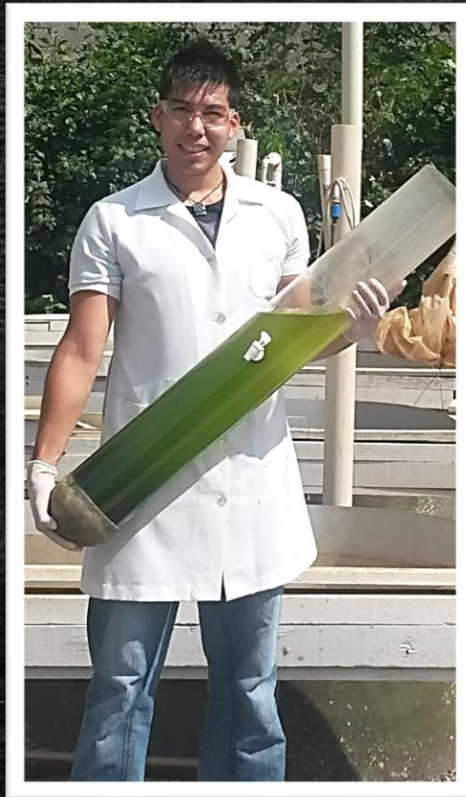
Sobre mim



- Engenharia Ambiental
 - Universidade Federal de Viçosa



Sobre mim



- Engenharia Ambiental
 - Universidade Federal de Viçosa
- Marketing Digital
 - Stone Co.
- Head de Marketing
 - Associação Brasileira de Fintechs
- Business Intelligence
 - Neoassist



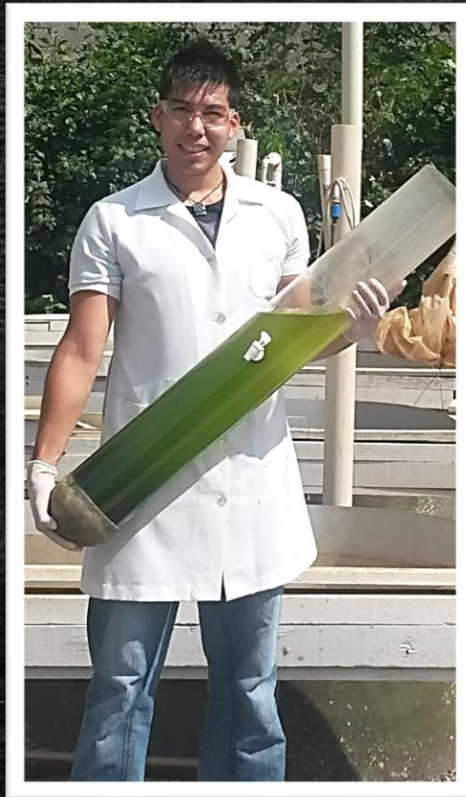
Sobre mim



- Cientista de Dados
 - Itaú Unibanco
- Espec. Data Science
 - Instituto Tecnológico de Aeronáutica
- Espec. Big Data Analytics
 - Fundação Instituto de Administração
- Professor de Data Science
 - Let's Code Academy



Sobre mim



- Cientista de Dados Sênior
 - Gol Linhas Aéreas
- Professor de Data Science
 - Trading com Dados
- Cientista de Dados Sênior
 - Diversos segmentos
- MSc. Statistics and Data Science
 - Katholieke Universiteit Leuven

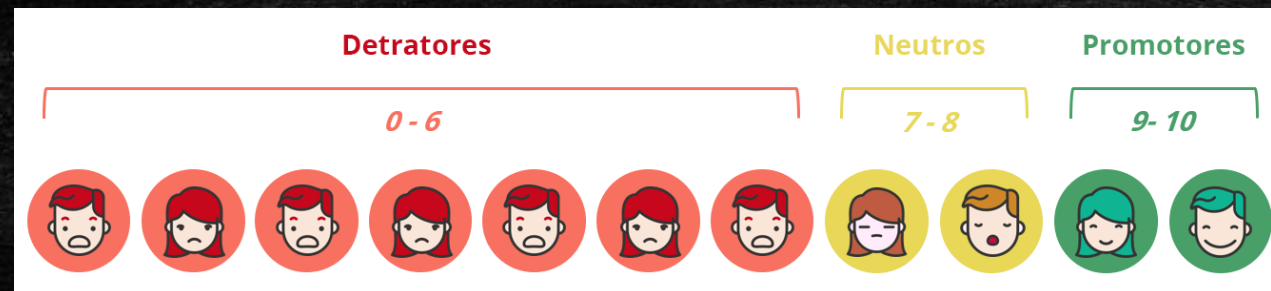


KU LEUVEN



Contexto e problema de negócio

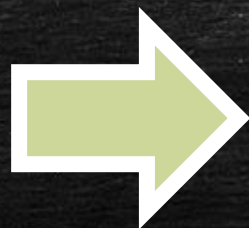
- Contexto: área de satisfação ao cliente de empresa de telecomunicações
- Problema: Alto número de detratores de acordo com pesquisa de NPS (net promoter score)
- Solução: Machine Learning para prever tipos de detratores para cada problema utilizando texto de pesquisas de NPS



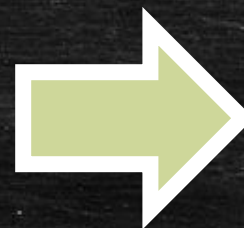
Ideia da solução



Comentário
do usuário



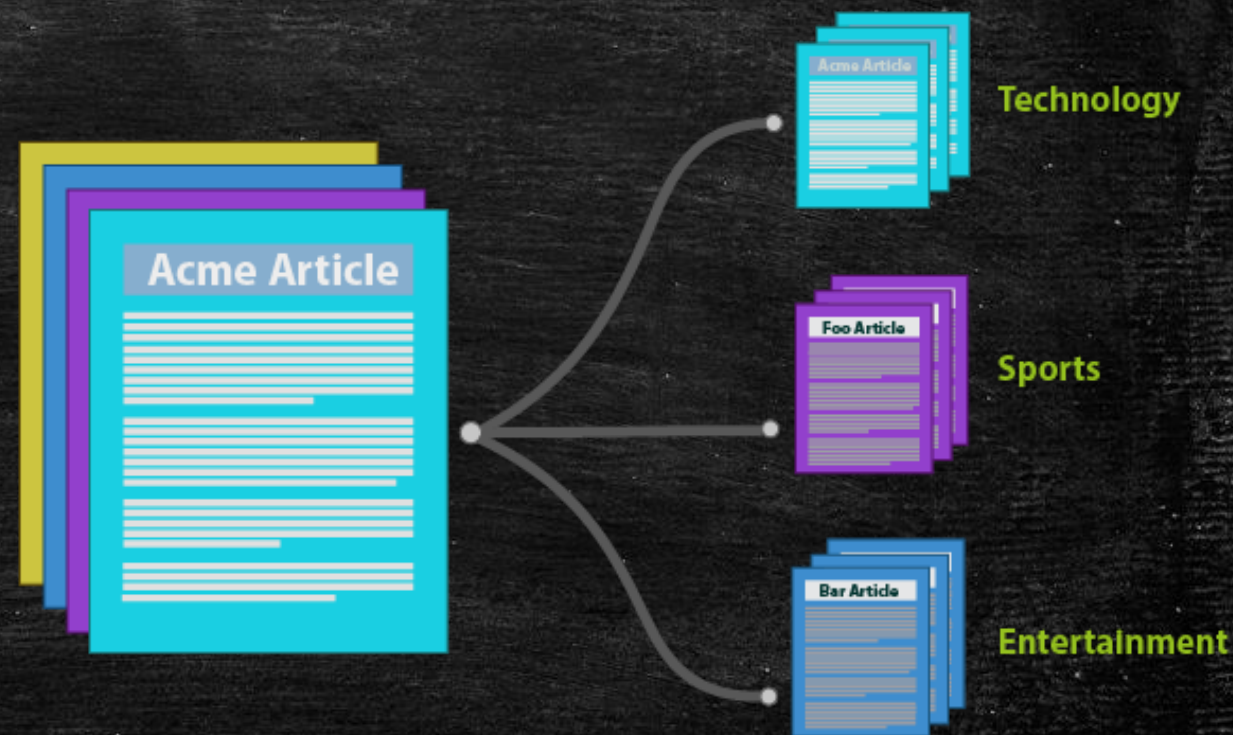
Algoritmo de
Machine Learning



Predição de
detratores

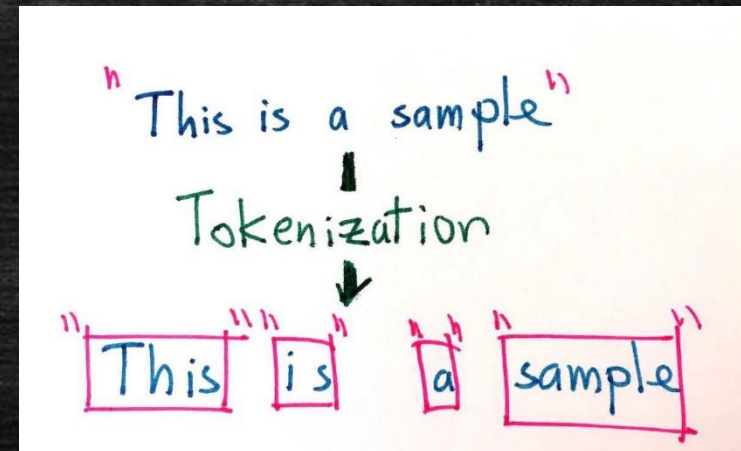
NLP – Natural Language Processing

- Textos de pesquisas são dados não estruturados
- Podemos adotar diversas abordagens



NLP – Natural Language Processing

- Tokenização:
 - O texto é separado em palavras e alocado em cada coluna
 - O número de linhas permanece o mesmo do dataset



	good	movie	not	a	did	like
good movie	1	1	0	0	0	0
not a good movie	1	1	1	1	0	0
did not like	0	0	1	0	1	1

NLP – Natural Language Processing

- BoW – Bag of Words
- Vantagens:
 - Simples de entender
 - Fácil de implementar
- Desvantagens:
 - Pode aumentar muito a dimensão de variáveis explanatórias
 - Não infere significado semântico entre as palavras

The Bag of Words Representation

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...

NLP – Natural Language Processing

- TF-IDF – Term Frequency – Inverse Document Frequency
 - Medida utilizada para comparar número de vezes que uma palavra aparece no mesmo documento e em outros documentos
 - Ideia é dar um peso:
 - Positivo para palavras que aparecem muitas vezes no mesmo documento
 - Negativo para palavras que aparecem muitas vezes em vários documentos

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

tf_{ij} = number of occurrences of i in j
 df_i = number of documents containing i
 N = total number of documents

Better BOW

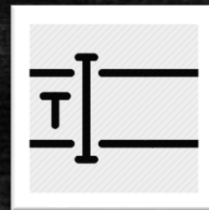
- Replace counters with TF-IDF
- Normalize the result row-wise (divide by L_2 -norm)

	good movie	movie	did not	...
good movie	0.17	0.17	0	...
not a good movie	0.17	0.17	0	...
did not like	0	0	0.47	...

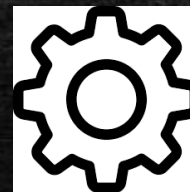
Pipeline de Machine Learning



Remoção de
palavras sem
significância



Tokenização e
TF-IDF



Tuning de
hiperparâmetros

Hands on!!!





Nome: Milton Ossamu

Contatos: [linkedin.com.br/in/ossamum](https://www.linkedin.com.br/in/ossamum)