

# SOLUÇÕES DE AUTOMATIZAÇÃO DE PROCESSAMENTO DE TEXTO

Demonstrar aplicação de Processamento de  
Linguagem Natural (NLP)  
no contexto de três soluções desenvolvidas no  
TCU.

Alessandra de Queiroz Requena Garrido ([alessandraqr@tcu.gov.br](mailto:alessandraqr@tcu.gov.br))

Alexandre Vaz Roriz ([alexandrevr@tcu.gov.br](mailto:alexandrevr@tcu.gov.br))

Marcus Vinicius Borela de Castro ([borela@tcu.gov.br](mailto:borela@tcu.gov.br))

# PLANO DE VÔO

Quebra-gelo

Contextualização

Apresentação das Soluções

ClaDoP (classificador de tipo de documento)

Verificador de Erros Materiais em Acórdãos

Zello (chatbot)



# MAS O QUE É DATA SCIENCE?

Demonstrar aplicação de Processamento de Linguagem Natural (NLP)  
no contexto de três soluções desenvolvidas no TCU



# O QUE É DATA SCIENCE?

Google search results for "data science x statistics venn diagram".

Imagens

o que quis dizer: data science vs statistics venn diagram

The Essential Data Science Venn Diagram - knugggets.com

Battle of the Data Science Venn Diagrams - knugggets.com

Battle of the Data Science Venn Diagrams - knugggets.com

Battle of the Data Science Venn Diagrams - knugggets.com

The Data Scientist Venn Diagram - rpubs.com

Battle of the Data Science Venn Diagrams - knugggets.com

Understanding Different Combinations of Skills - dimensionless.in

The Data Science Venn Diagram - hub.packtpub.com

The Data Scientist Venn Diagram - knugggets.com

Driving the Success of Data Science Solutions: Skills, Roles and Responsibilities - knugggets.com

The Essential Data Science Venn Diagram - knugggets.com

Learning Statistics Online for Data Science - dimensionless.in

Battle of the Data Science Venn Diagrams - knugggets.com

Introduction To Data Science - Learn - learn.co

Venn diagram of variation partitioning of C. b... - researchgate.net

What is data science? What does it require? - quora.com

Five Misconceptions about Data Science - rpubs.com

RPubs - ModernDive: Statistical Inference - rpubs.com

Battle of the Data Science Venn Diagrams - knugggets.com

What is Data Science? - becomingdatascientist.com

What are the educational qualifications for Data Science? - quora.com

How to Become a Data Scientist | Springboard - springboard.com

Data science?21 - slideshare.net

How to Become a Data Scientist - Springboard - springboard.com

Driving the Success of Data Science Solutions: Skills, Roles and Responsibilities - knugggets.com

How to Become a Data Scientist - Springboard - springboard.com

# CONTEXTUALIZAÇÃO – DATA SCIENCE



Ciência de Dados (Data Science)

Estatística

Área interdisciplinar voltada para o estudo e a análise de dados, estruturados ou não, que visa extrair conhecimento (ou *insights*) e apoiar tomadas de decisão



Ciência aplicada a dados

# CONTEXTUALIZAÇÃO (VISÃO DE TÉCNICAS)

Análise de Dados  
Ciência de Dados aplicada

Processo de explorar um conjunto de dados, extraindo ou ajudando a evidenciar padrões e auxiliando na descoberta de conhecimento

Mineração de Dados  
Ciência de Dados (Data Science)  
Subárea da Análise de Dados

# CONTEXTUALIZAÇÃO (VISÃO DE TÉCNICAS)

Texto é um tipo de dado (não estruturado)

Ciência de Dados (Data Science)

Análise de Dados

Técnicas e rotinas relativas a Linguagem

NLP (Processamento de Linguagem Natural)

Processo de explorar um **texto** extraído ou ajudando a evidenciar padrões e auxiliando na descoberta de conhecimento

Mineração de Dados  
Text Mining



# CONTEXTUALIZAÇÃO (VISÃO DE TÉCNICAS)

As técnicas em NLP podem ser agrupadas em 3 níveis:

Rotinas básicas (usadas em pré-processamento)

- Tokenização
- Stemização
- Busca por expressões regulares
- ...

Tarefas de apoio (“blocos de lego”)

- Identificação de conjuntos frequentes
- NER (Named Entity Recognition)
- Dependency Parsing
- ...

Objetivos autônomos

- Classificação de tipo de documento
- Tradução de texto
- Sumarizador de texto
- ...

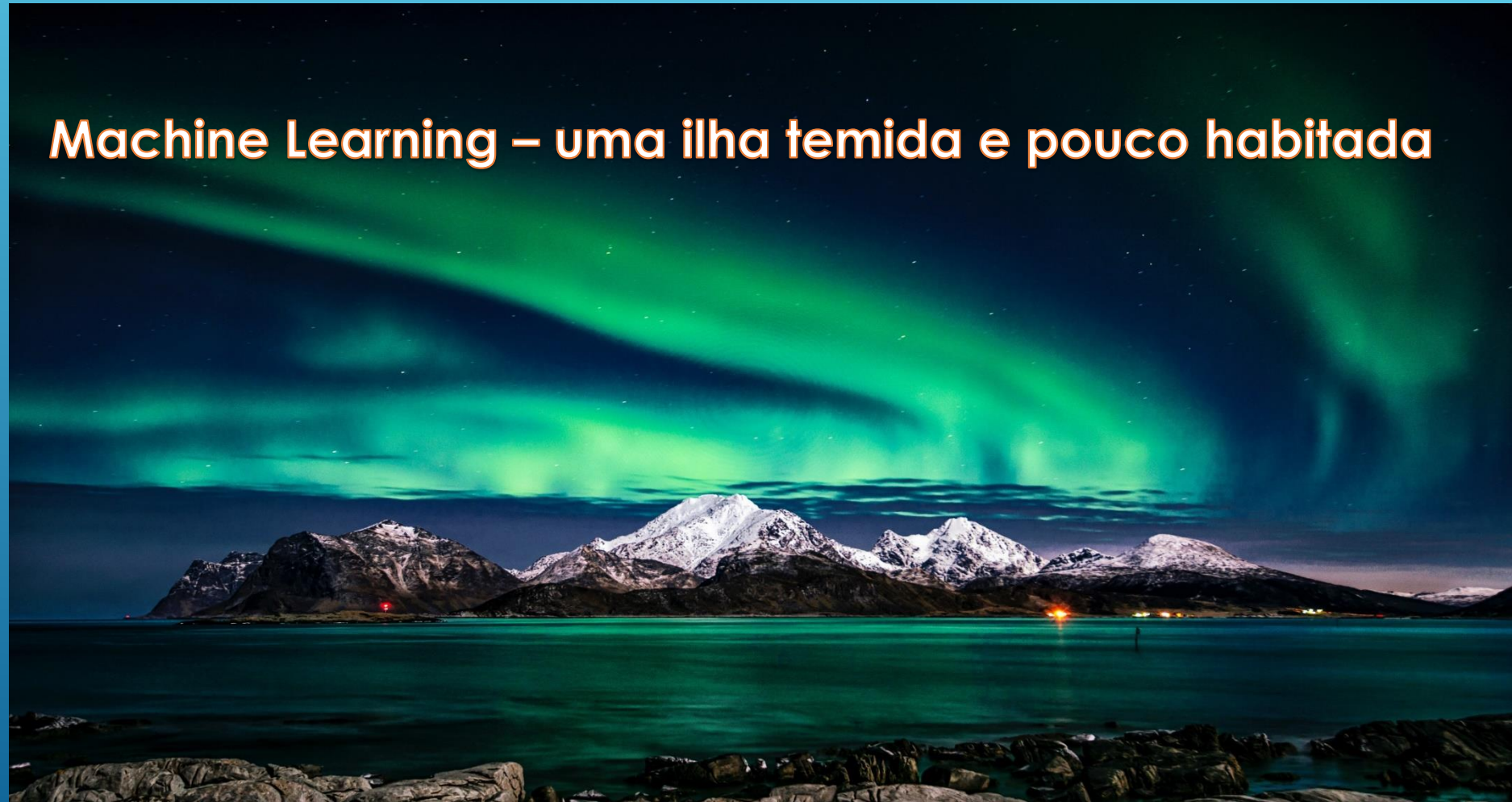
Fonte interessante (papers recentes por tarefas/objetivos): <http://nlpprogress.com/>





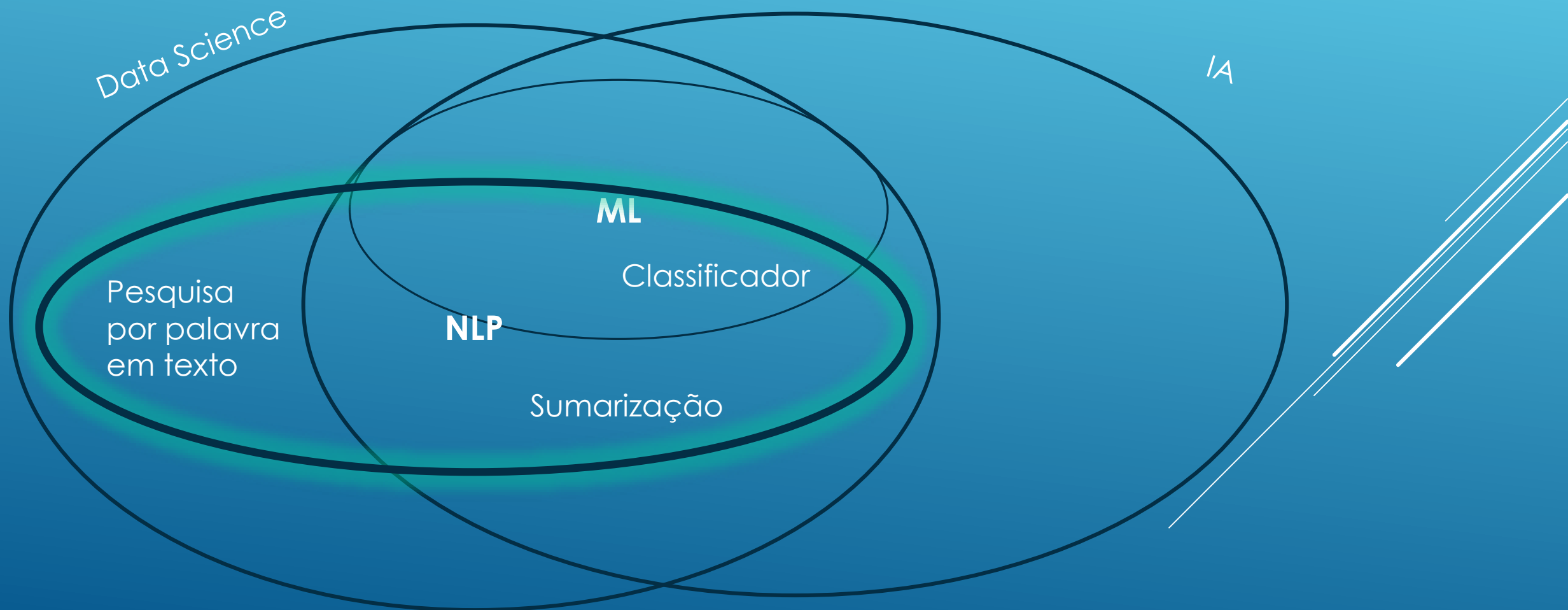
# DÚVIDA QUE NÃO QUER CALAR...

E NLP: é IA ou Data Science? Ou é Machine Learning (ML) ?

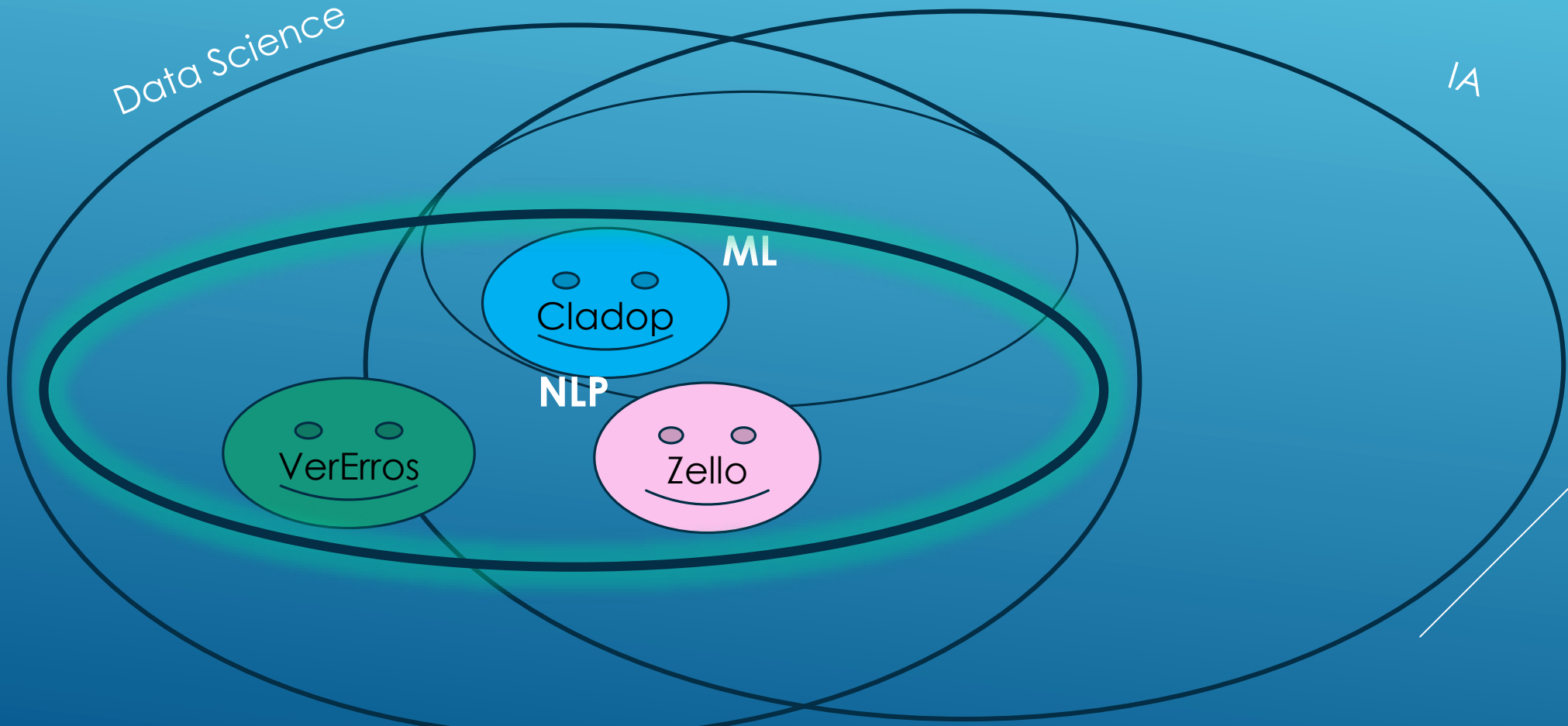


# DÚVIDA QUE NÃO QUER CALAR...

E NLP: é IA ou Data Science? Ou é Machine Learning (ML) ?



# CONTEXTO DAS SOLUÇÕES A SEREM APRESENTADAS (DE NLP)



# PLANO DE VÔO

Quebra-gelo

Contextualização

Apresentação das Soluções

ClaDoP (classificador de tipo de documento)

Verificador de Erros Materiais em Acórdãos

Zello (chatbot)