SAS® Club

Der Business Analytics Club für SAS User
8. September 2022 | Wien

8. September 2022

14:00 - 14:10 Uhr	Begrüßung / Intro / News Gerhard Svolba, SAS
14:10 - 14:40 Uhr	Erstellen einer benutzerdefinierten Funktionsprozedur mit PROC FCMP in SAS zur Berechnung des multilateralen Verbraucherpreisindex Adam Tardos, Statistik Austria
14:40 - 14:55 Uhr	Self-Service Analytics in an industrial use case Manfred Kügel, SAS
14:55 - 15:10 Uhr	KI in Aktion - Bilderkennung und Natural-Language-Processing mit SAS Gerhard Svolba, SAS
15:10 - 15:25 Uhr	Python, R und SAS: Getrennte Sprachen, gemeinsame Prozessflüsse Gerhard Svolba, SAS
15:25 - 15:35 Uhr	Pause
15:35 - 15:50 Uhr	SAS und DevOps Phillip Manschek, SAS
15:50 - 16:05 Uhr	SAS Studio – der Nachfolger des SAS Enterprise Guide und mehr Phillip Manschek, SAS
16:05 - 16:20 Uhr	Tipps und Tricks: Wie Sie mit SAS Procedures animierte Graphiken erstellen und Bewegung in Ihre Darstellungen bekommen können Gernot Engel, SAS
16:20 - 16:35 Uhr	SAS Information Catalog, schnell die richtigen Daten finden & verstehen Gernot Engel, SAS
16:35 - 16:50 Uhr	SAS Support-Änderung in 2025 Phillip Manschek, SAS
16:50 Uhr	Q&A

KI in Aktion - Bilderkennung und Natural-Language-Processing mit SAS

Gerhard Svolba, SAS Austria
(credits to
Ian McCallum IIASA, Laxenburg,
Jon Walker, SAS HC North Carlolina)

Links

- https://www.sas.com/de_at/data-for-good/rainforest.html
- https://developer.sas.com/home.html
- https://developer.sas.com/guides/dlpy.html
- https://github.com/sassoftware/python-dlpy
- **IIASA, SAS. (2022) Crowd-driven deep learning tracks Amazon deforestation. *In prep,***
  - ***Ian McCallum, International Institute for Applied Systems Analysis, Laxenburg, Austria,***
  - ***Jon Walker, SAS Campus Drive, Cary, NC 27513, United States***

Statistik, Machine & Deep Learning

Forecasting, Optimierung

Daten Visualisierung

Model Deployment

Daten Management

Decision Management

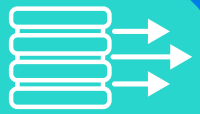
Natural Language Processing

Computer Vision

Künstliche Intelligenz

ist die Disziplin, Systeme zu trainieren, um Abläufe der menschlichen Tätigkeit durch LERNEN und AUTOMATISIERUNG zu emulieren.

Lernen aus Erfahrungen/ Daten



Anpassen an neue Fakten

Automatisierung des Prozesses

sas

Analyse der Abholzung des Amazonas Regenwalds auf Basis der automatischen Klassifikation von Satellitenbildern

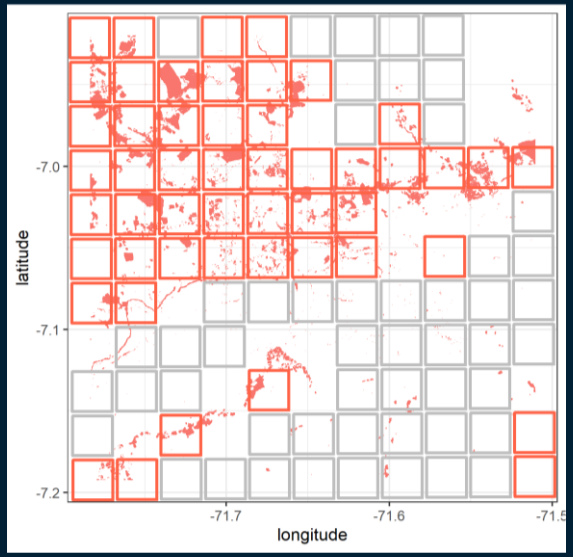
(Kooperation mit der IIASA, International Institute for Applied System Analysis)

https://www.sas.com/de_at/data-for-good/rainforest.html

Background and Project Goal

- Amazon Rainforest 5,5 Mio km^2 (larger that EU)
- Deforestation:
  - 29000km^2/year in 2003
  - 6000km^2/year in 2014,
  - now raising again

- Project Goal:
  Train a computer vision model to automatically detect areas where deforestation is taking place



Überblick über den Modellierungsablauf

Satellite Images

Perform Expert Rating (Subset)

Train a first Model

Model 1 “Expert”

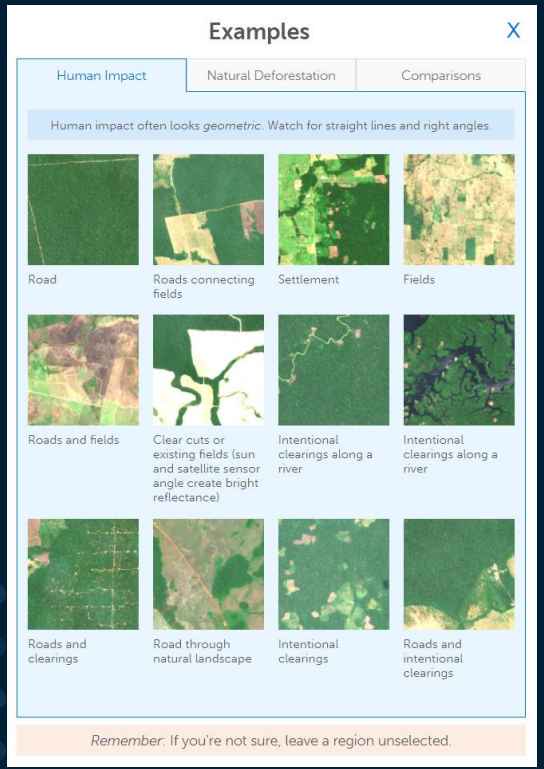
Validate Model Accuracy

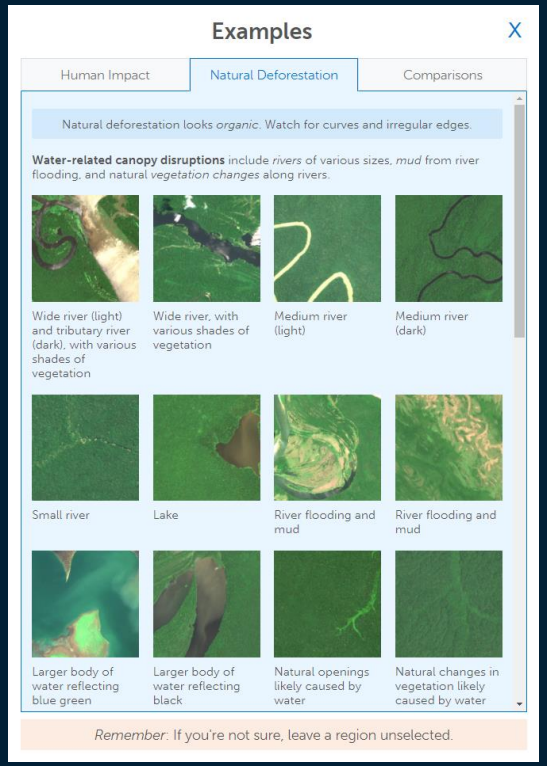
Provide Data for Crowd Sourcing

Images

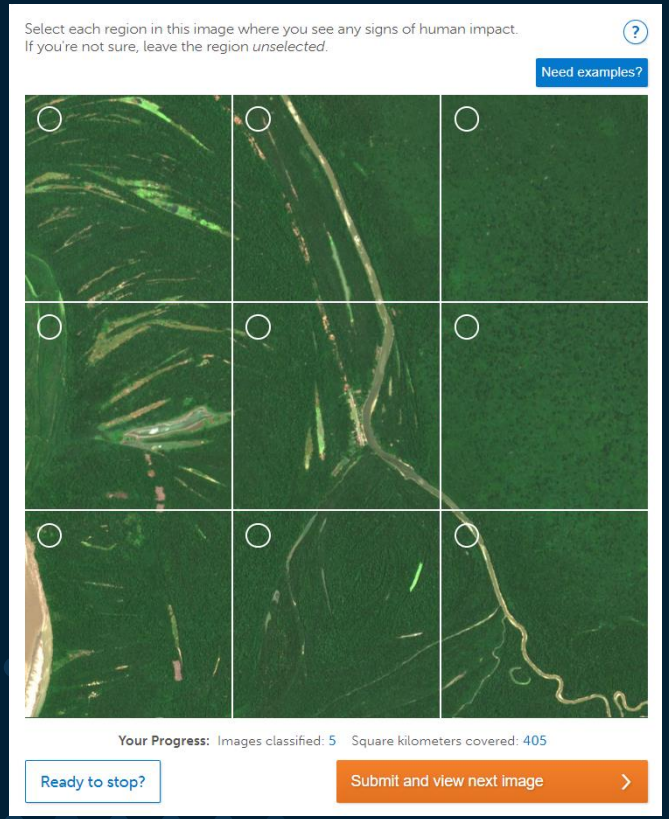
Crowd Sourcing (GatherIQ)

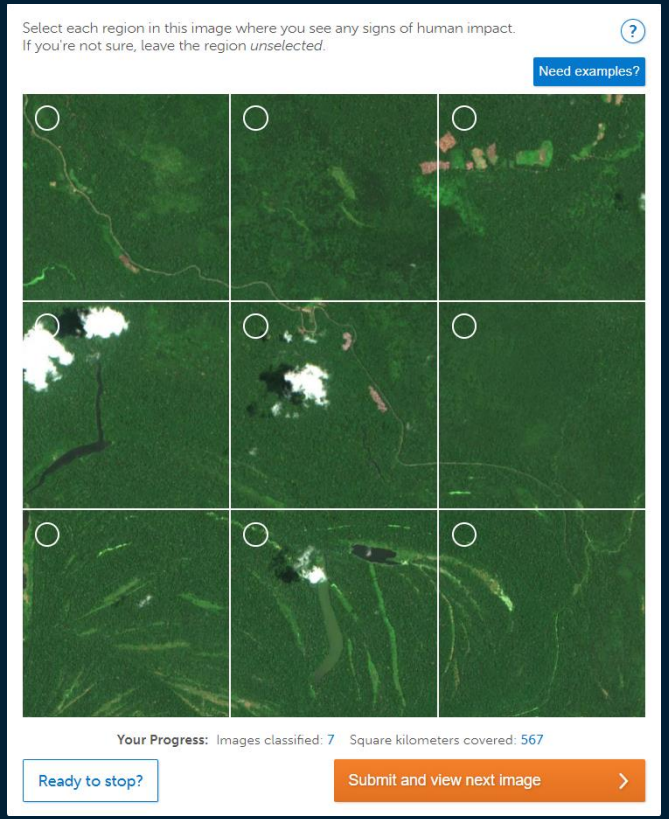
Menschlicher Engriff oder natürliche Entwaldung?





Beispiele für die Anwendung der Klassifikations-App



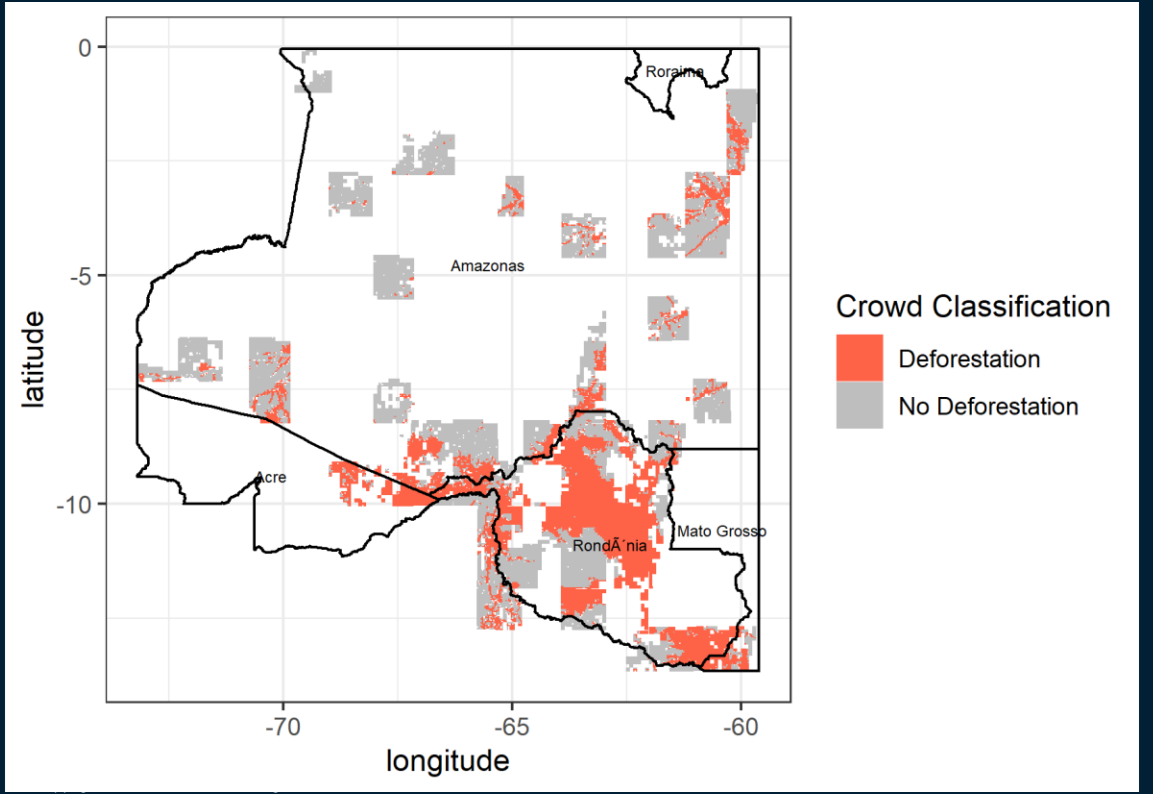


Crowd Sourcing

- Studies show: perform equally well as experts
- Accuracy increasing when consensus or majority voting is used
- Only few rainforest deforestation pictures exist in ImageNet, CIFAR-10, CIFAR-100. And: not in 10m resolution as from Sentinel-2 satellite
- Deforestation project: 6 months, 5500 active participants from 96 countries, 389.988 km2
- Used cloud free samples from sentinel-2 satellite images
- Consensus among the crowd: >80% for majority of pictures
- Agreement with expert review on a sample of 200 pictures: 88%

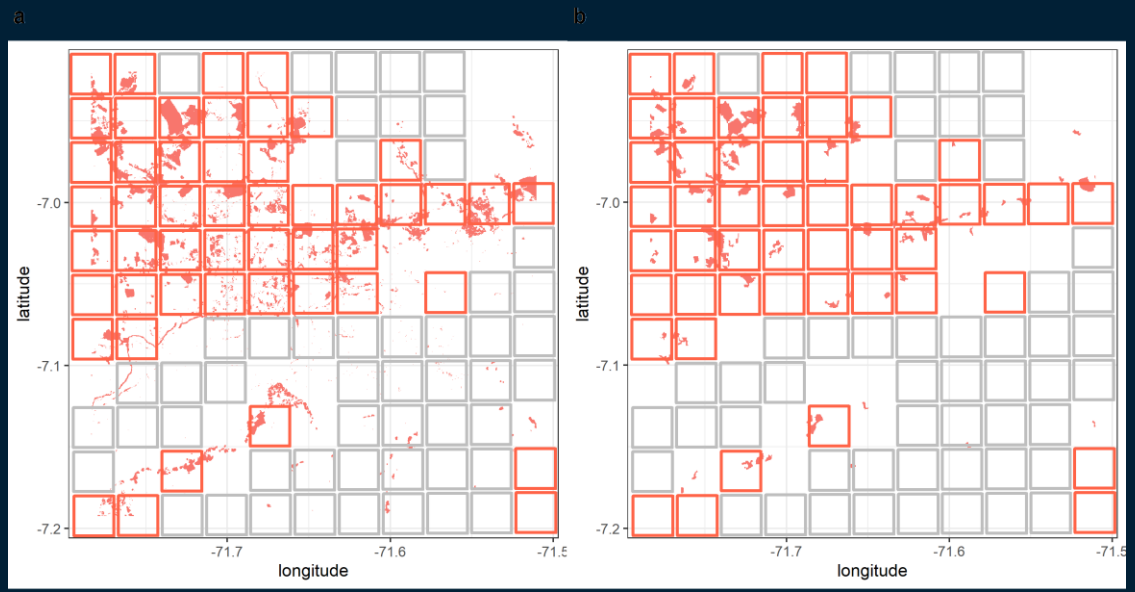
Results from Crowd Sourcing

Figure 1. Results of the crowdsourcing campaign over the Brazilian Amazon between June and November 2020. Map of the 390,000 km^2 (43,100 images) classified by the crowd as having either evidence of deforestation or no deforestation. Individual pixels represent a 3 x 3 km image.

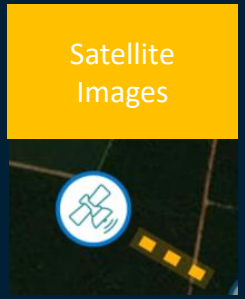


Comparing with existing products

- Comparison with existing products (PRODES and Maryland) difficult as they have differences in spatial details.
- After searching for >1km2 signs to compare with crowd results: 92% accuracy with PRODES and 89% with Maryland.



Überblick über den Modellierungsablauf



https://github.com/sassoftware/python-dlpy

README.md

DLPy - SAS Viya Deep Learning API for Python

An efficient way to apply deep learning methods to image, text, and audio dat

SAS VIYA 3.4 | PIP INSTALL SAS DLPY | PYTHON 3+

Overview

DLPy is a high-level Python library for the SAS Deep learning features available in SAS Viya. D
to provide an efficient way to apply deep learning methods to image, text, and audio data. D
created following the Keras APIs with a touch of PyTorch flavor.

What's Recently Added

- Text, audio, and time series support in addition to image
- New APIs for:
  - RNN based tasks: text classification, text generation, and sequence labeling
  - Object detection
  - Image segmentation
  - Time series processing and modeling

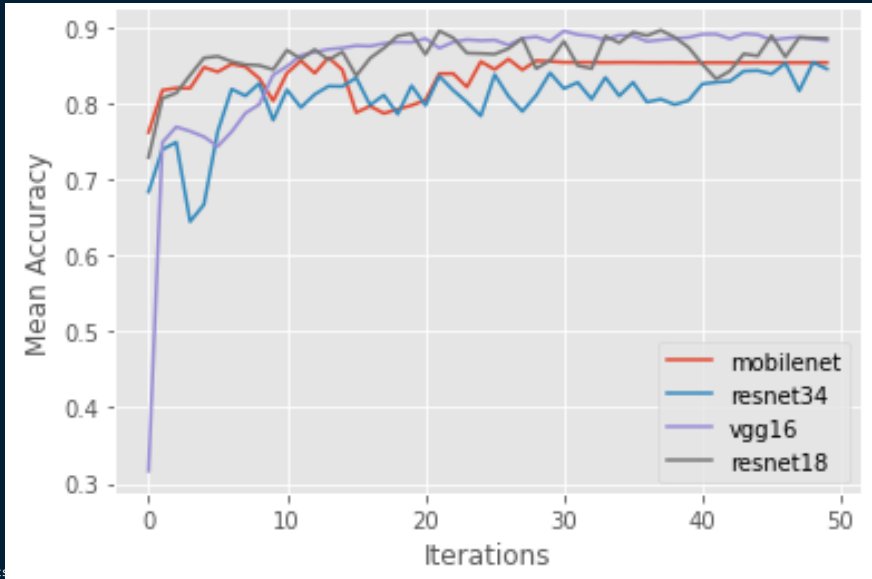
master · python-dlpy / examples / · Go to file · Add fi

ChipRobie updated audioTraining example runs end-to-end with new supplied data — 7a425c7 on Apr 29, 2021

..	
auto_encoder	Updated autoencoder example. (#319)
functional_model	Adding bulding_model.rst
image_captioning	Hide formatting tags in image captioning example. (#316)
image_classification	Update image classification with EfficientNet example. (#314)
image_denoising	New autoencoder denoising example. (#309)
image_embedding	Notebook Edits
keras_model_conversion	Merge pull request #293 from ChipRobie/DEEPLRN-245
keypoints_detection	Edit keypoints example
learning_rate_policy	Update learning rate example notebook. (#313)
misc	- adding mist folder to examples as well as an example of video proce...
multitask_learning	adding another example notebook
object_detection	Fast RCNN Soccer: update input dataset and model weights
onnx	Revised and clarified import ONNX model example. (#317)

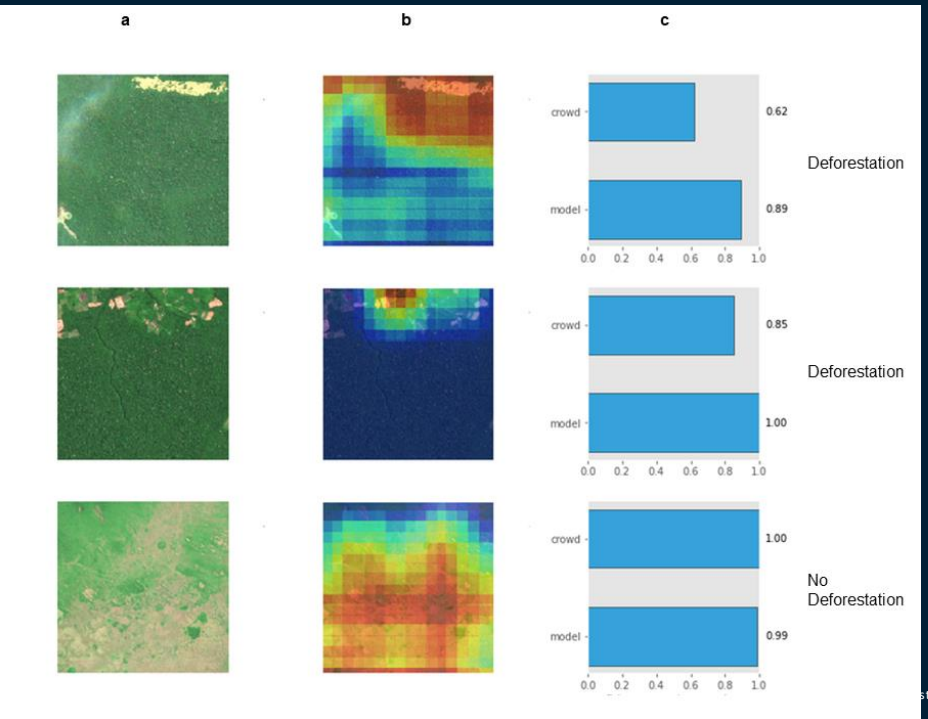
Deep Learning Methods

- 43000 images in the crowdsourced library
- 60/20/20 split for train, validation, test
- Models trained: VGG1629, ResNet18, ResNet3430, MobileNet31.
- ResNet18 model slightly outperformed all others



Example results and their interpretation

Test dataset | occlusion sensitivity | confidence of the crowd



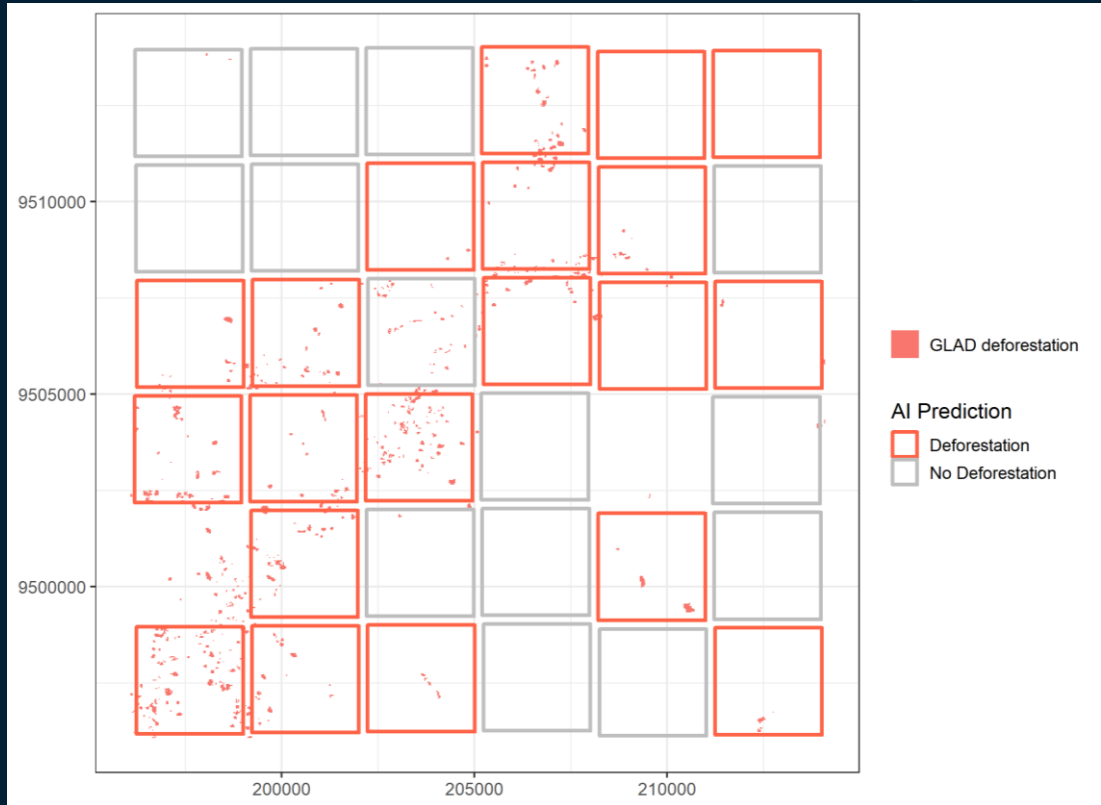
signs of deforestation in the image that trigger the activation layers, giving the model high confidence that human impact has occurred even though the crowd was less confident

signs of human impact at the top of the image trigger the activation layer accordingly

both the crowd and the model identify the canopy disturbance as non-human impact with high confidence

Validation of the AI Model

- 94,8 agreement with the crowd on 8774 images
- 100 (unseen) sentinel-2 satellite images: comparing results between the AI model and the GLAD-S2 deforestation dataset



Data Selection and Preprocessing

- Crowd sourcing is based on the visible spectrum (RGB). Sentinel-2 contains 13 optical bands of which we used three to create RGB images (Blue (~493nm), Green (560nm), and Red (~665nm)) with a 10 m resolution.
- Data pre-processing:
  - 5% cloud filter has been used
  - Picture with excessive coloud coverage or with majority of missing data have been removed
- Computer vision model is not restricted to visible wavelengths.
- Models that use the near-infrared spectrum or the actual computed indices, such as NDVI, may be able to more accurately distinguish between natural deforestation, water features, and otherwise disturbed or developed land.

Überblick über den Modellierungsablauf

Satellite Images

Perform Expert Rating (Subset)

Train a first Model

Model 1 “Expert”

Validate Model Accuracy

Analyze the effect of having additional data (labels)

Provide Data for Crowd Sourcing

(Continuously) train the augmented model

Validate Model Accuracy

Model 2 “Crowd”

Images

Crowd Sourcing (GatherIQ)

Labels

Longitudinal Analysis (Stability+ Forecasting)

“Fresh” Data (Periodically)

Brazilian Public Images

Other public Images

Spectral Preprocessing

Scoring of new images

Validate Model Accuracy

Crowd vs. Model: Comparing the classifications

- established as an alert system complementary to the existing monitoring systems in place
- address the limited uptake of existing monitoring systems into actionable change or policy developments, through inclusion of the global civil community
- crowd appears to perform better in terms of classifying rivers and roads (water can appear in different colors, reflection)
- Additional filtering on image data could be applied
- Image size for crowd sourcing: 3 x 3 km for visual quality and identification of large areas. However less precise about the exact location

Text-Analytik
(Natural Language Processing)

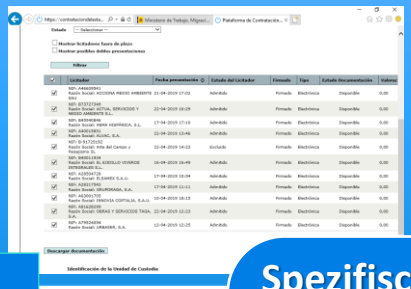
Einsatz von Technologie, um die menschlichen Handlungen des Lesens, Organisierens und Quantifizierens von Freiform-Texten auf sinnvolle Weise zu skalieren.

„Smart Document“ Fallbeispiel

Erhöhung des Service-Levels
und der Zufriedenheit der Bürger/Innen

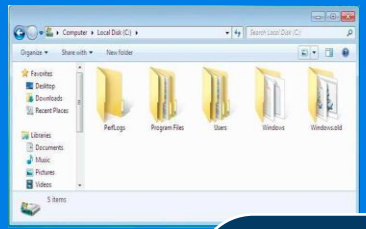
Die Einbettung des „SMART Document“ Systems

Spanische eProcurement Plattform



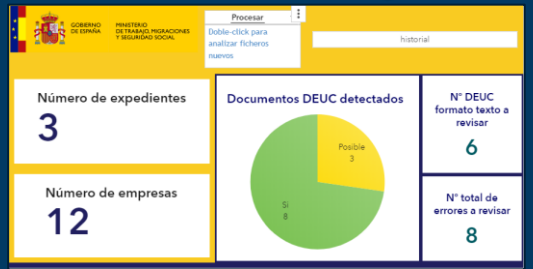
Jurist entscheidet sich für das Herunterladen des Ausschreibungsdossiers mit allen vorgestellten Unternehmen

Spezifische Verzeichnisse



Speichert das Dossier in einem Verzeichnis

SAS Anwendung



Starten der automatischen Analyse für ein neues Gebot

Automatische Erkennung von Anomalien in Dokumenten und Anwendung der fachlichen Regeln

Analyse der Dokumentation des Bieters

Überblick über das Ökosystem

Prozess-Automatisierung

Dokument Extraktion	Dokumenten-Klassifikation	Entitäten Generierung	Daten-Aufbereitung	Ergebnisse Berichte

Ausgangspunkt

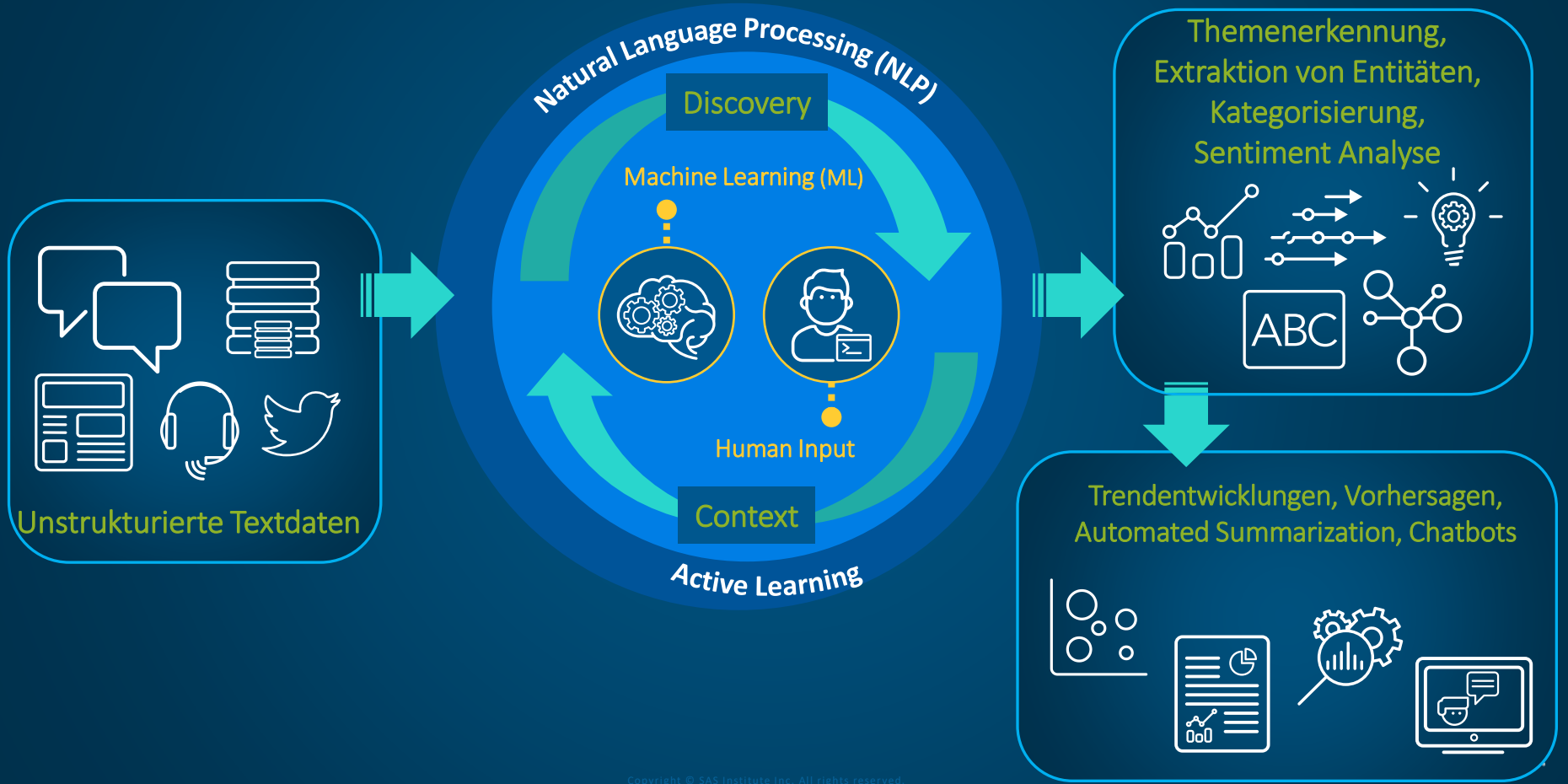
Download aus dem ESPD

Ziel

Nombre empresa	CIF	fileName	DEUC
KPMG AUDITORES , S . L .	[redacted]	DEUC.pdf	Si
DELOITTE , S . L .	[redacted]	DEUC_Fdo.pdf	Si
CREATIVIDAD Y TECNOLOGÍA S . A .	[redacted]	1DEUC9386CYTSA.pdf	Si

Tabelle mit der entsprechenden Dokumentation pro Gebot

NLP und Maschinelles Lernen



SAS Visual Text Analytics in SAS Model Studio