Continental
The Future in Motion

Data Science Challenge

Tire Identification

HACKATHON
DATA SCIENCE

Filipa Castro (filipa.castro@conti.de)
Joana Silva (joana.raquel.silva@conti.de)

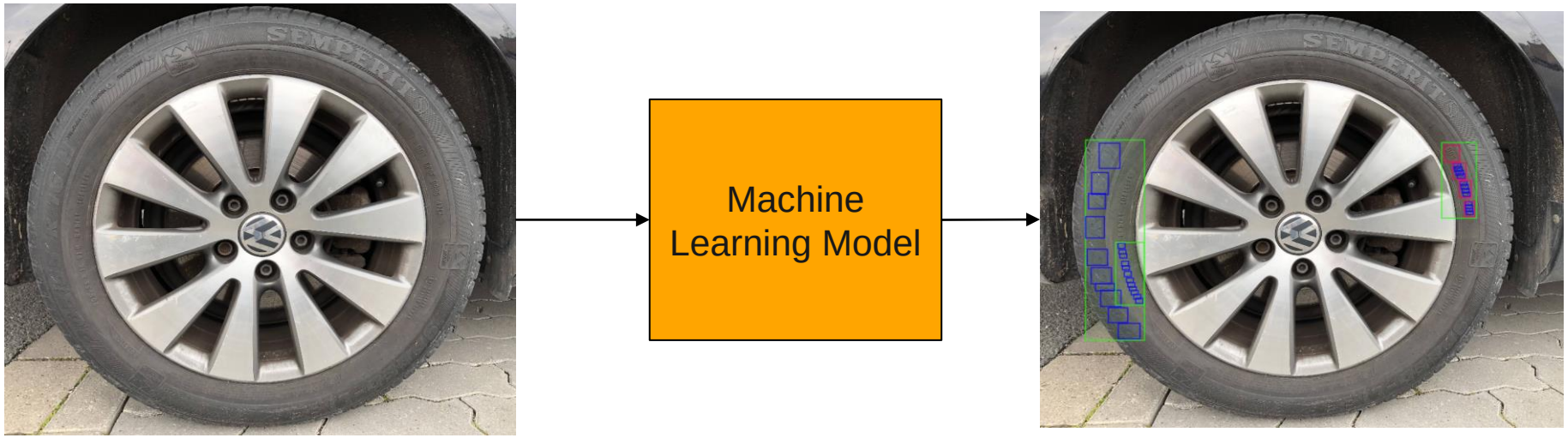
APPLIED ANALYTICS + AI

DATATTACK

Table of contents

Data Science Challenge

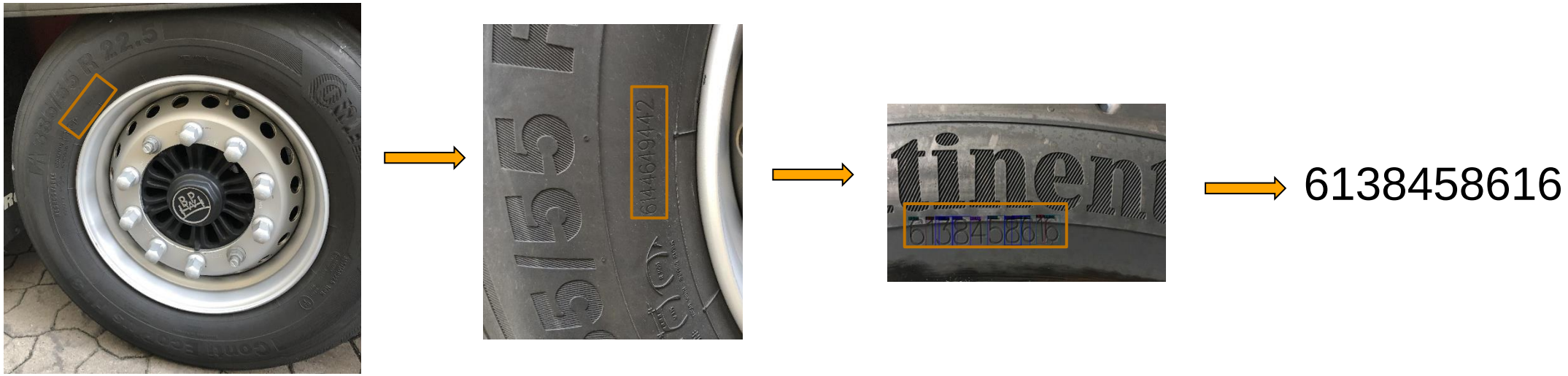
› **The goal** is to create an image recognition engine that, given an image of a tire sidewall, allows the automatic reading of character information (markings).



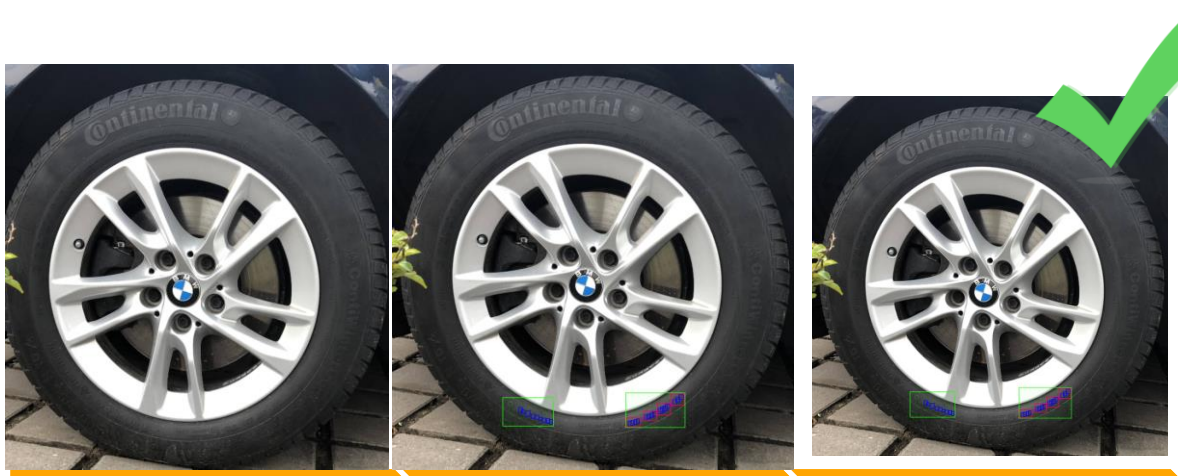
Dimension: 195R55 86H
DOT: CPO FDRH 54717
…

Data Science Challenge

- ✓ **Use case:** to develop an Optical Character Recognition (OCR) of serial tire identification number (serial number) on Commercial Specialty Tires (CST).



How to create AI using image recognition



Datattack

Image Acquisition → Image Annotation → Data Quality Check → Algorithm Training → Algorithm Validation → Algorithm Deployment

Timeline of Hackathon

Analysis of image recognition dataset

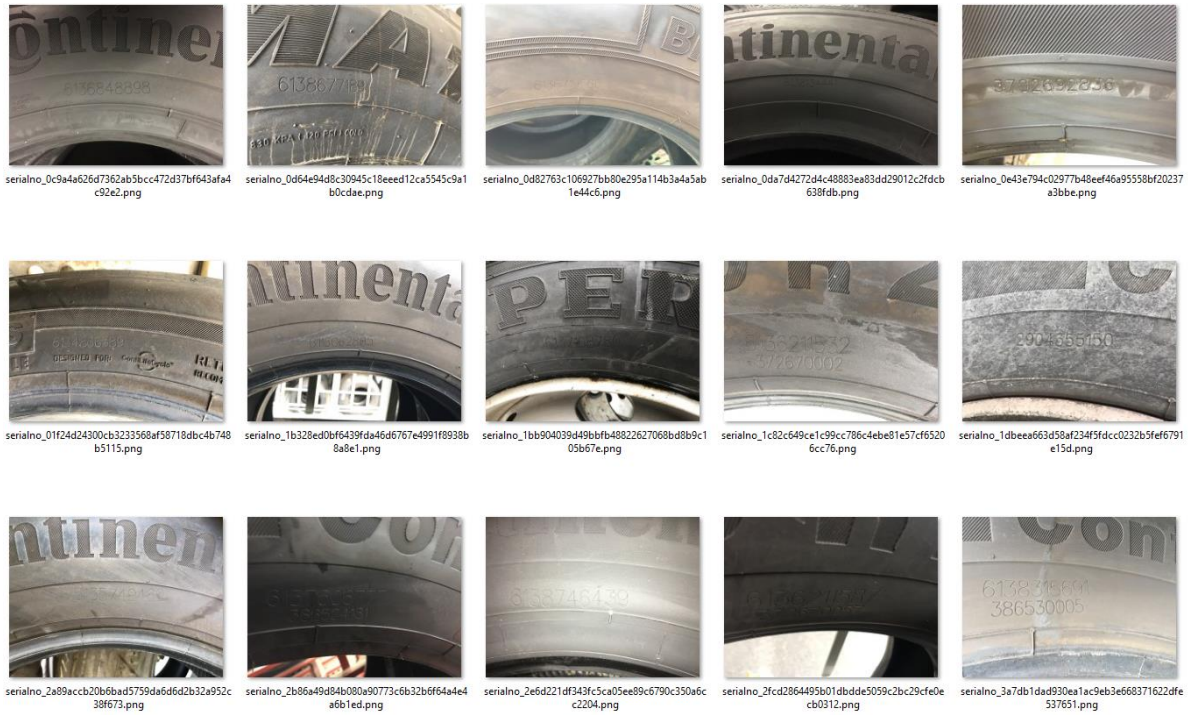
Sidewall markings recognition using image algorithms to detect the serial number of the CST tires

Prepare training and test sets to develop a robust model

Submit final predictions for a new test set that will be provided by the end of the competition (Sunday morning)

Dataset

- Training set (80%)
  - We have 295 images of Commercial Vehicle Tires (CVT), cropped in the serial number region. The annotations are in COCO format.
    - labels_train.json: 295 train set annotations
- Testing set (20%)
  - The test data will be made available for 15 minutes on Sunday morning. The teams have 15 minutes to run the inference and submit the predictions in the expected csv format.
    - labels_test.json: 74 test set annotations




```
{
    "id": 1295,
    "image_id": 104,
    "category_id": 3,
    "area": 12144.0,
    "bbox": [
        423.0,
        824.0,
        88.0,
        138.0
    ],
    "iscrowd": 0,
    "attributes": {
        "batch_cvat": "empty",
        "train_test": "train",
        "brand": "TBD",
        "orientation": null
    }
},
```

```
{
    "id": 1296,
    "image_id": 104,
    "category_id": 7,
    "area": 11288.0,
    "bbox": [
        510.0,
        820.0,
        83.0,
        136.0
    ],
    "iscrowd": 0,
    "attributes": {
        "batch_cvat": "empty",
        "train_test": "train",
        "brand": "TBD",
        "orientation": null
    }
},
```

Data vizualization

- A script will be provided to launch Voxel FiftyOne app for data vizualization at https://github.com/NIAEFEUP/datattack

Filter annotations by character

Evaluation Procedure

- The **Levenshtein distance** is a string metric for measuring the difference between two strings. It is calculated as the minimum number of single-character edits necessary to transform one string into another.

- Participants should provide a csv file for the test set in the following format:

	image_name	string_1	string_2	string_1_prediction	string_2_prediction	ls_distance_1	ls_distance_2	ls_distance_avg
0	serialno_e2801d5c6408262ec75a0e661496172467fc38bc.png	6133548541		8574212507	665909561	9		9
1	serialno_97b3a0c993f6bc08652f407bcd60541025701fce.png	6133548541		4066669589	575421672	9		9
2	serialno_5da8c43b37c67d37a65f642147bcfe6c03f055a3.png	6134202576		9392758237	103903242	8		8
3	serialno_43c2644a27af2cabed686ca012cd095fcf2d843e.png	6133548574		3197478441	630975254	8		8
4	serialno_f83f17ad7fd0380e2a8d0adb305bdcffb5f8bc62.png	6129844192		3395393632	765739039	9		9
5	serialno_99b455d0fadbd7d167d30a3a5d8a6e622ff8a2cf.png	6115535543		7384826006	424932615	10		10
6	serialno_aed1f6ea4a289e521db611cd320cbbca593ac940.png	6115535543		5985077972	402980042	9		9
7	serialno_ab6980a9fb988cb771af75faff1e73a030099b81.png	6130769635		8596189794	378800657	9		9
8	serialno_6725e617ca16792ab0c5c7d1e4389bb88f2a5595.png	6129419586		4602774453	435419566	8		8
9	serialno_14774fd83267809bcab241c40258b2d79a8a2683.png	3701427804		1938758491	516087145	10		10
10	serialno_b980b255720ec18f407eab6ef21aa8047dcc34c4.png	3702044013		2271915740	295396260	9		9
11	serialno_3a761604a2e64b0882792c639cb486b9185f3d8c.png	6136309131	372672052	9848306324	835473601	8	10	9

- A script for exporting predictions in this format, given a directory of test images, is given at https://github.com/NIAEFEUP/datattack

Evaluation Criteria

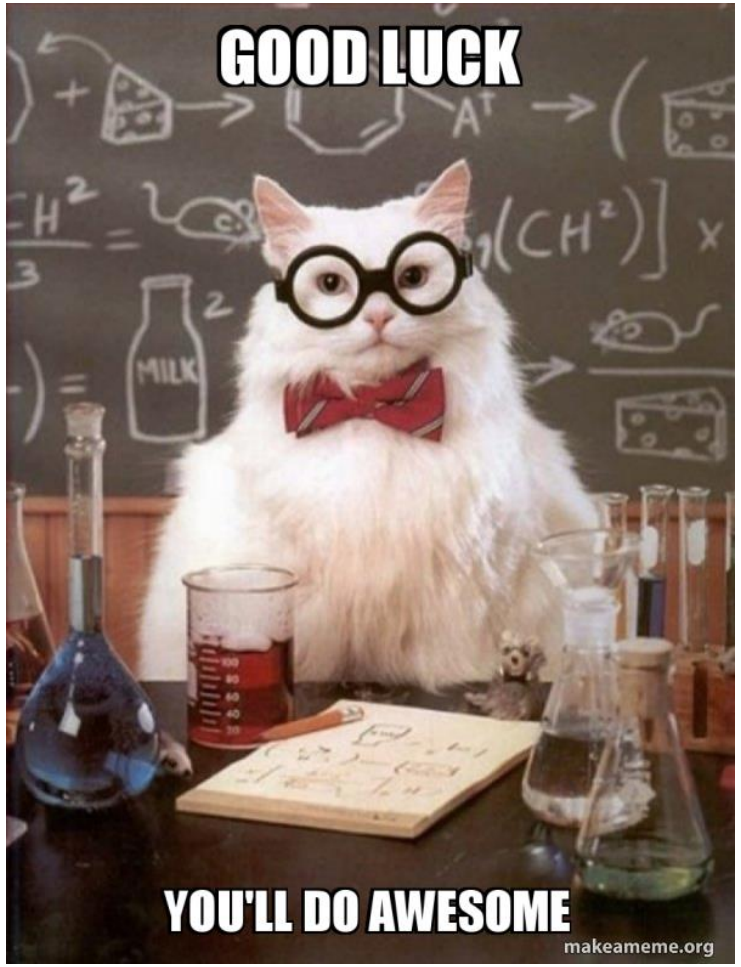
- Teams will also be evaluated with additional criteria:
  - Pitch for the presentation
  - Technical solution approach and performance
  - New collected and annotated data
    - Acquire more CST tire images (how many images were acquired and annotated) and evaluate how the model works for these new images.

- Participants are expected to submit a zip folder with the name [TEAM_NAME].zip
  - New_data: containing additional images collected and annotations in COCO format
  - Code: containing all the scripts used for the solution
  - Pitch presentation slides

- <u>Do NOT distribute the data & DELETE it after the challenge</u>

Useful information

- Use Google Colab to develop and train the model (GPU available for free):
  - https://colab.research.google.com/?utm_source=scs-index
- Create a new google account for the team, as it will make it easy to work together with Colab and to mount the drive folder with the dataset on the Colab
- Use the Discord channel to Q&A about the challenge:
  - https://discord.com/channels/968655877529796688/968655877529796691
- Use ImgLab to annotate the new acquired dataset
  - https://imglab.in/
- You can use Voxel FiftyOne to see the dataset and annotations:
  - https://voxel51.com/



1st Place	300€
2nd Place	200€
3rd Place	100€