

Taak "Machine Learning"

Project "Musti"

**HO
GENT**

Taak: project “Musti”: situering

- Musti is een huiskat (*Felis silvestris catus*) die via een kattenluik vrij in en uit het huis van haar baasjes kan gaan. Een chip garandeert dat geen vreemde katten binnen kunnen.
- Haar baasjes weten graag op elk moment of Musti binnen of buiten is, ook als ze zelf niet thuis zijn.



Taak: project “Musti”: situering (vervolg)

- Ze hebben daarvoor een camera met bewegingsdetectie geïnstalleerd die bij elke beweging 6 foto's (met tussentijd van één seconde) neemt: net vóór de bewegingstrigger en de beweging zelf.
- Deze bewegingsdetectie werkt echter op basis van het percentage gewijzigde pixels en wordt dus ook geactiveerd bij bv. een gewijzigde lichtinval.

Taak: project “Musti”: situering (vervolg)

- Natuurlijk herkent de camera ook Musti niet: een insect voor de lens of een voorbijlopende bewoner (of is het een indringer?) activeert ook de camera.
- Dat kan beter m.b.v. Machine Learning!

Taak: project “Musti”: trainingsdata

- De baasjes zijn bezig met het verzamelen van de foto's van de bewegingsdetectie en deze te classificeren in drie klassen:
 - “aanwezig”: Musti staat op het beeld maar gaat niet naar buiten
 - “buiten”: Musti is op weg naar buiten
 - “niets”: Musti staat niet op het beeld

Je vindt deze trainingsdata (die regelmatig wordt aangevuld) op: <https://bit.ly/ProjectMusti>

Taak: project “Musti”: voorbeelden trainingsdata



Aanwezig



Buiten

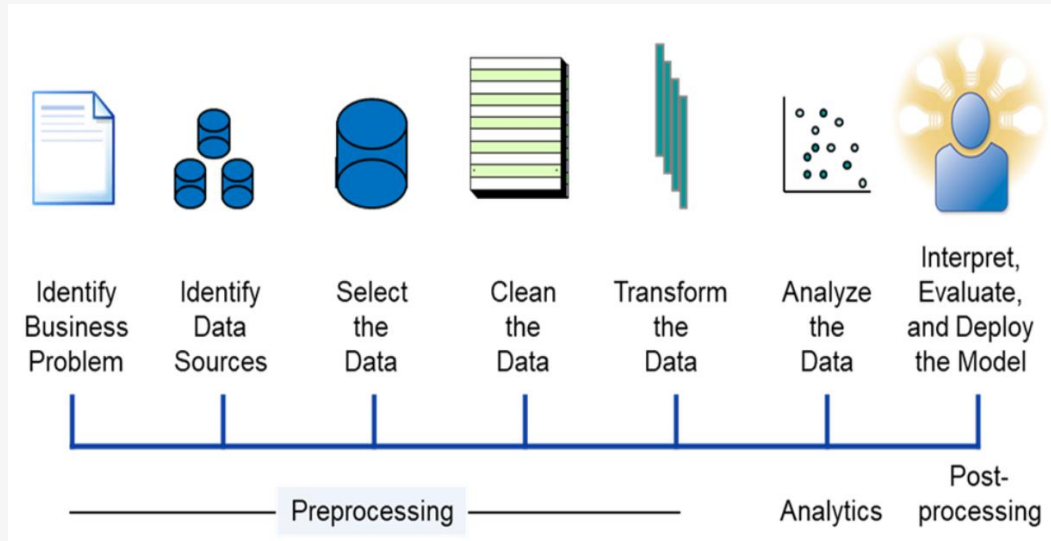


Niets

Taak: project “Musti”: taak

- Stel een classificatiemodel op voor de trainingsdata.
 - Kies een gemotiveerde maat voor de nauwkeurigheid van je model.
 - Kies een optimaal classificatiemodel.
- Stel de confusion matrix op voor de 3 klassen en bespreek.
- Maak een Python-applicatie met als functionaliteit
 - Op een website (één html-file met foto) wordt de huidige status van Musti gepubliceerd: “binnen” of “buiten”, met tijdstip van de foto waarop deze status vastgesteld is en de foto zelf.
 - De URL van de website is parametrizeerbaar.
 - Bepaal de betrouwbaarheid van de status en verwerp statussen onder een bepaalde (experimenteel bepaalde) drempelwaarde.
 - De mogelijkheid tot hertrainen van het model bij aangevulde trainingsdata.

Houd rekening met: het data mining proces



± 80-90 % van de totale projecttijd gaat naar preprocessing wegens:

- Bad veracity!
- Data en business probleem begrijpen is meestal niet eenvoudig.
- Algoritmes zelf zijn off-the-shelf beschikbaar (meestal geen PhD nodig)
- 80 % transpiratie, 20 % inspiratie

Taak: project “Musti”: evaluatie

- U werkt in groepen van 5 (inschrijven via Chamilo).
- Presentatie van project in week 13.
- Eén à twee vragen op schriftelijk examen specifiek over uw project.
- Samen (presentatie + examenvragen) 25 % van de evaluatie in eerste zittijd.
- Geen tweede examenkans → punten worden overgenomen.

Oefeningen (zie Chapter 1)

- Welk type machine learning zal je gebruiken volgens de verschillende typeringen?
- Welke uitdagingen zie je in dit project?