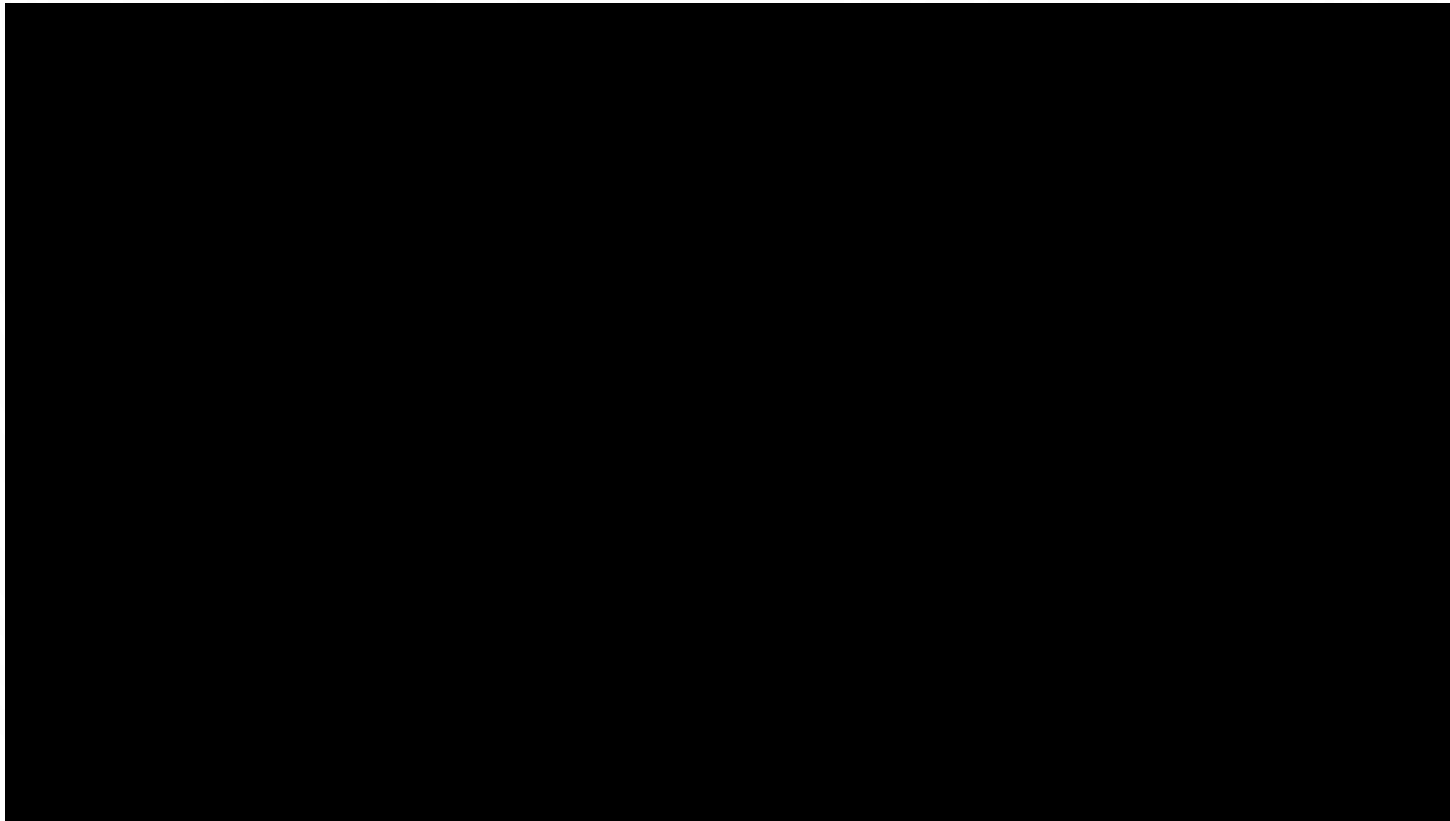


The background features a series of concentric circles in a light gray color, centered on the slide. Overlaid on these circles are stylized circuit board traces in a light blue color. These traces are located in the corners of the slide, with some ending in small circles, resembling nodes or components in a network or data flow diagram.

ANOMALY DETECTION

JENS BAETENS

WAT ZIJN OUTLIERS?



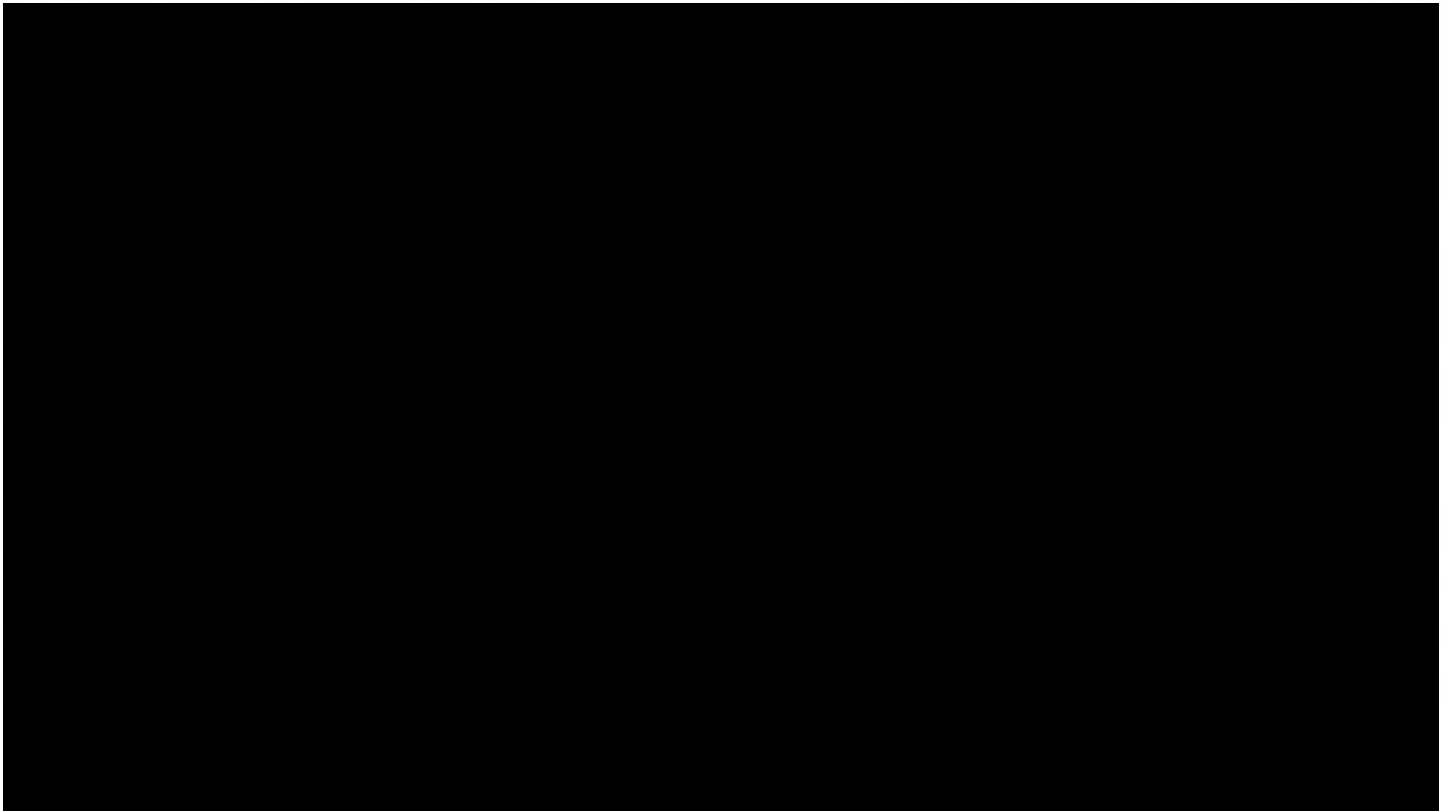
DOEL VAN OUTLIER DETECTION

Outliers = punten die zich niet gedragen als andere datapunten

Opsporen van problemen: medisch, mechanisch, fraude, hacking, ...

Vaak toegepast op niet-gelabelde data omdat niet alle problemen op voorhand vastgelegd kunnen worden en ze per definitie zeldzaam zijn.

VERWIJDEREN VAN OUTLIERS



VERWIJDEREN VAN OUTLIERS

Problemen met deze methode:

- Opnieuw trainen moeilijk in real-time situaties

Kan wel gebruikt worden om outliers te verwijderen in de trainingsdata maar niet in de testdata of na in gebruik name

Andere algoritmes:

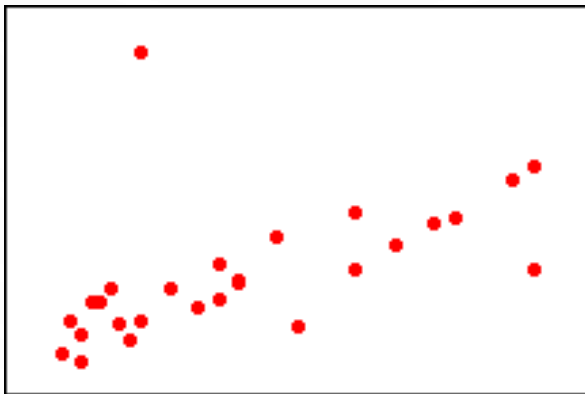
- Isolation Forests
- Local Outlier Factor
- One-Class-Svm
- Neurale Netwerken

ISOLATION FORESTS

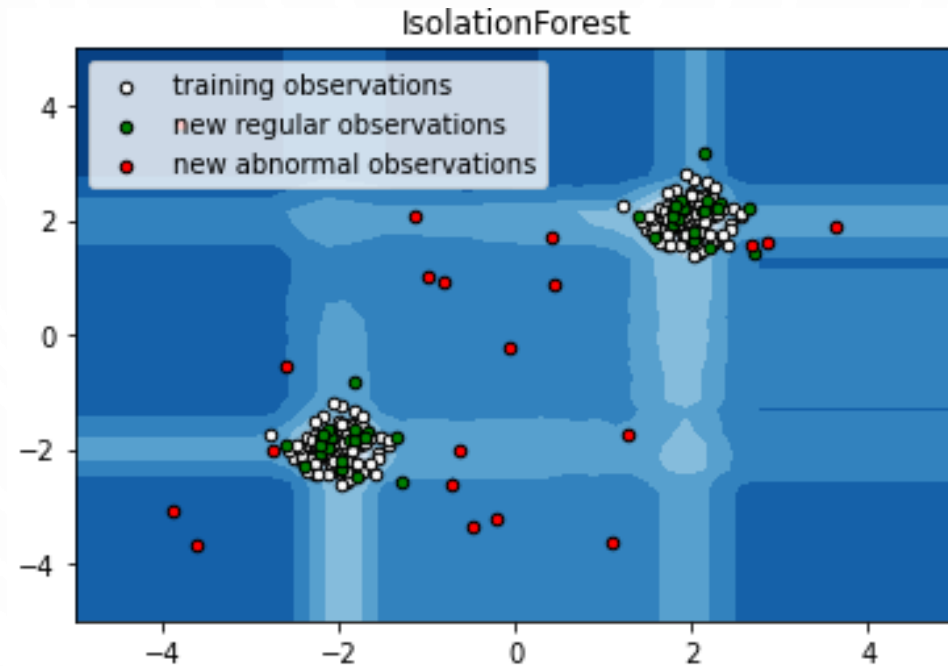
Ensemble techniek gebaseerd op random forest

Datapunten willekeurig verdelen tot elk punt alleen in een eindknoop zit.

Hoe korter het pad, hoe groter de kans op een outlier



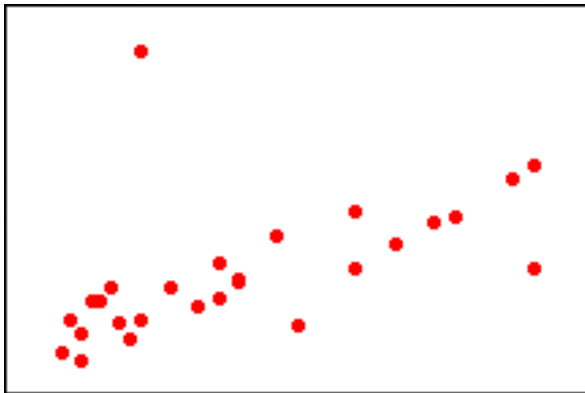
ISOLATION FORESTS



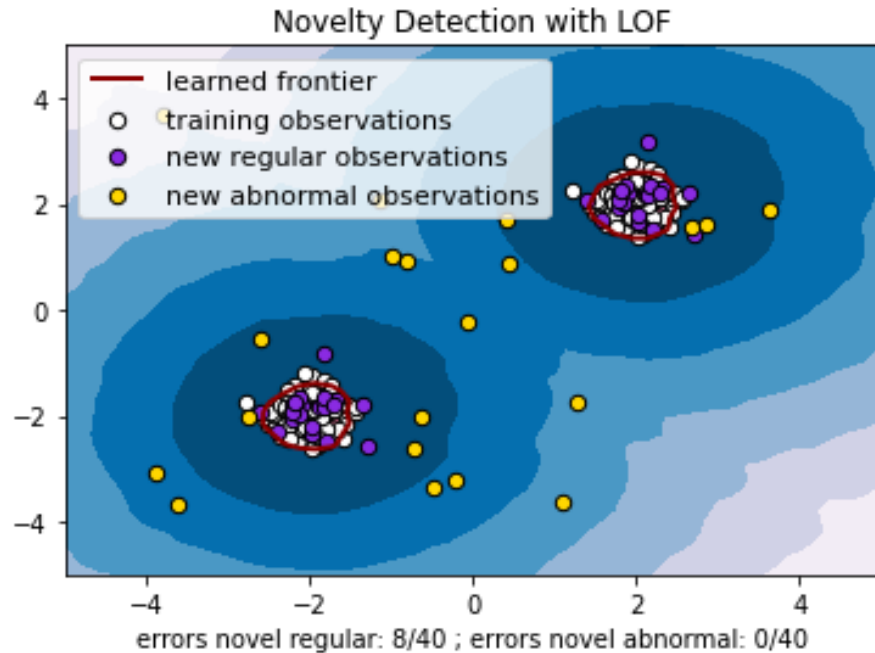
LOCAL OUTLIER FACTOR

Gebaseerd op K-Nearest Neighbours

Afstand tot deze burens bepaald voor kans in te schatten dat het een outlier is



LOCAL OUTLIER FACTOR



ONE CLASS SVM

Gebaseerd op SVM

SVM scheidt twee klassen met een hyperplane

One-Class SVM dat hypersfeer bepaald dat alle punten omvat

- Punten op de marge zijn de support vectors
- Mogelijkheid om een aantal “fouten” toe te laten
- Kernel om complexere sferen te bepalen

ONE CLASS SVM

