

SYV ÅRS ULYKKE

Kollisions data analyse med AI og ML

Erik Nielsen, Data Science Lead

Johan Hartnack, SYSTRA Danmark

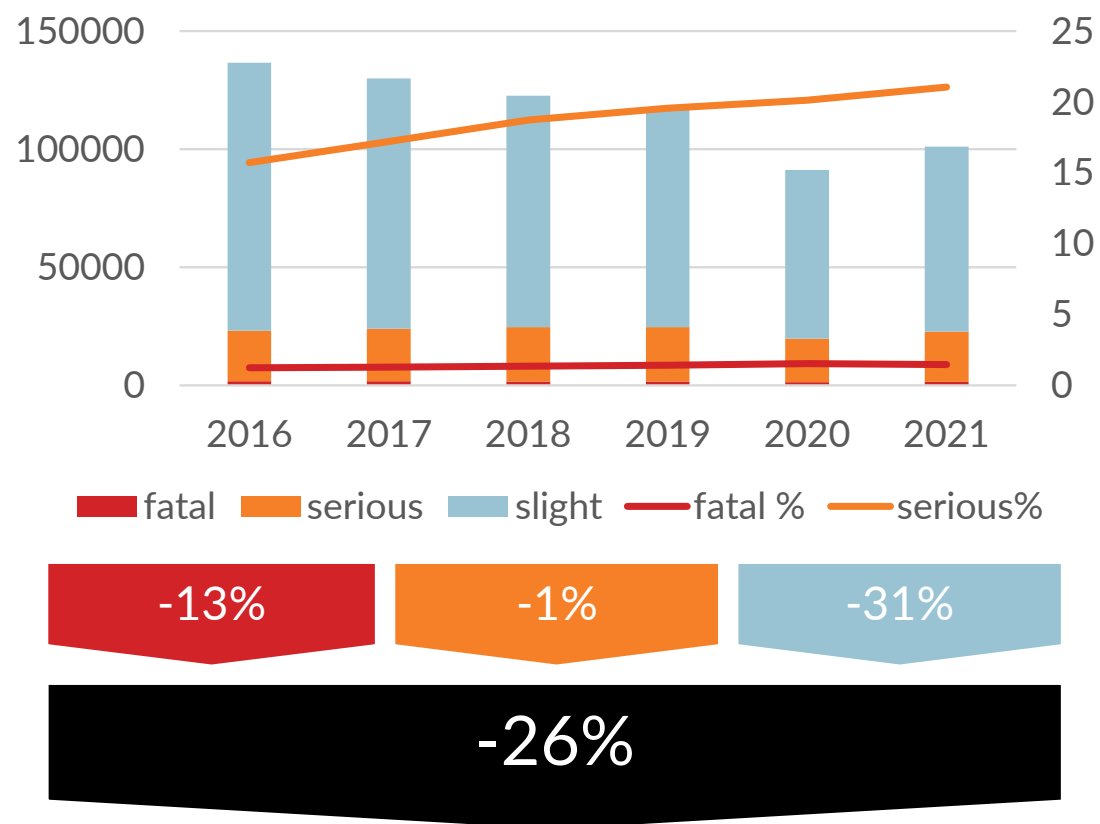
SYSTRA



Syv års ulykke

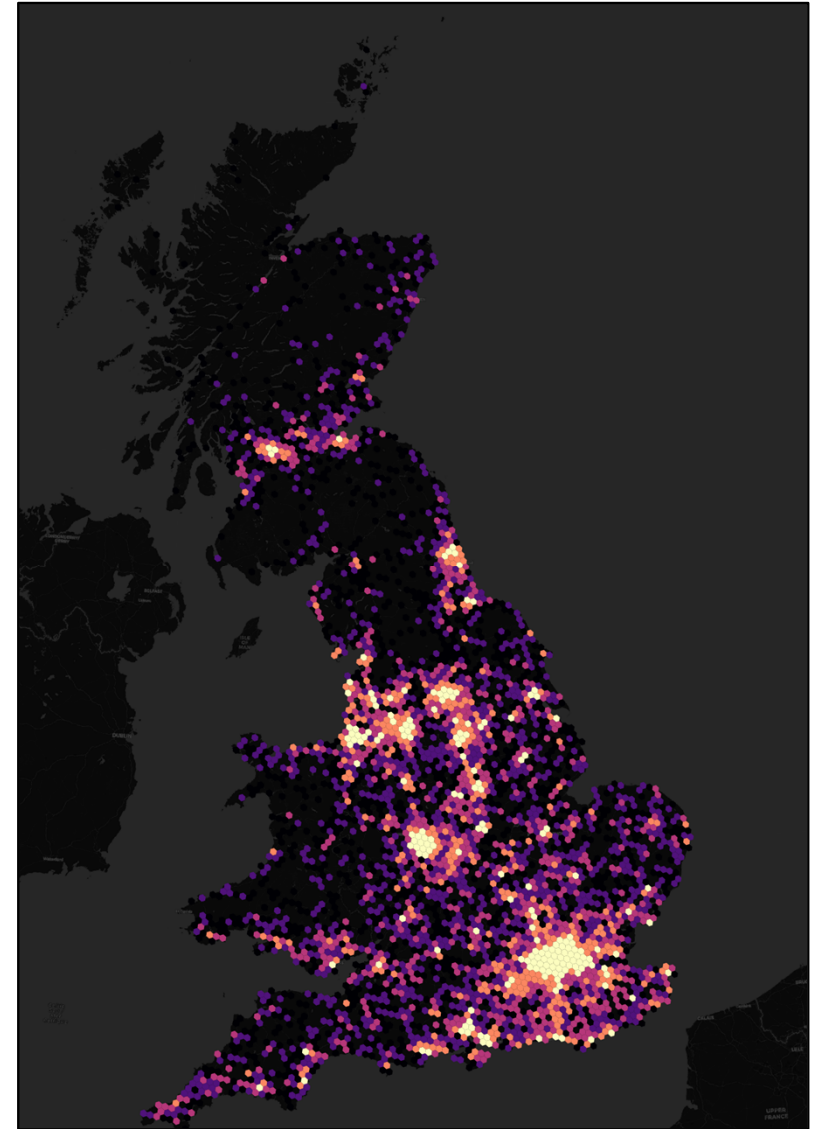
- Baggrund
 - Omkring 700.000 kollisioner rapporteret
 - Med lidt over 900.000 tilskadekomne
 - 10.000+ døde
 - 140.000+ svært kvæstede.
- Fokus på trafiksikkerhed
 - Store personlige og samfundsøkonomiske konsekvenser
 - Kontinuerlig indsats til reduktion af uheld i Storbritannien
- Hvordan kan vi bedre forstå og afværge de alvorligste kollisioner?

Kollisioner med personskaade



Baggrund

- SYSTRA Ltd har ved hjælp af **let tilgængelige** datasæt søgt at finde omstændigheder og årsagsfaktorer, der fører til dødsfald og alvorlige skader på vejene i Storbritannien.
 - Datakilder
 - Identificering, gap-analyse, data hygiejne
 - Data analyse
 - (Geo)statistik og visualisering
 - Data fletning
- Machine Learning



Datakilder

Primær datakilde er STATS19

- Statistikker om trafikssikkerhed baseret på ulykker rapporteret til politiet
- Åbent datasæt, opdateret to gange om året
- Indeholder et væld af information om hændelser:
 - Kollisiones alvor
 - Tilskadekomne
 - Type af involverede trafikanter/køretøjer
 - Vej- og vejrforhold

The image shows a detailed form for reporting an accident. It includes sections for:

- Accident Reference (1.3)
- Time and Date (1.4, 1.7)
- Road Details (1.1, 1.2)
- Reporting Officer (1.11)
- Vehicle and Casualty Counts (1.5, 1.6)
- Road Type (1.14)
- Junction Details (1.16)
- Weather (1.22)
- Junction Control (1.17)
- Light Conditions (1.23)
- Special Conditions at Site (1.24)
- Road Surface Condition (1.25)

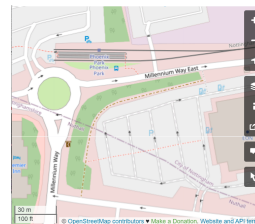
2.5 / 2.5a TYPE OF VEHICLE X	
Car	09
Taxi / Private hire car	08
Van - Goods vehicle 3.5 tonnes mgvw and under	19
Goods vehicle over 3.5 tonnes mgvw and under 7.5 tonnes mgvw	20
Goods vehicle 7.5 tonnes mgvw & over	21
M/cycle 50cc and under	02
M/cycle over 50cc and up to 125cc	03
M/cycle over 125cc and up to 500cc	04
Motorcycle over 500cc	05
Motorcycle - cc unknown	07
Electric Motorcycle	23
Pedal cycle	01
Bus or coach (17 or more passenger seats)	11
Minibus (8-16 passenger seats)	10
Agricultural vehicle (include diggers etc)	17
Ridden horse	16
Mobility scooter	22
Tram / Light rail	18
Other 1	00
vehicle 2	00
3	00
4	00

2.10 JUNCTION LOCATION OF VEHICLE X	
Not at or within 20m of junction	0
Approaching junction or waiting /parked at junction approach	1
Cleared junction or waiting /parked at junction exit	2
Leaving roundabout	3
Entering roundabout	4
Leaving main road	5
Entering main road	6
Entering from slip road	7
Mid junction- on roundabout or on main road	8

Sekundære kilder

- Ordnance Survey MasterMap
- OpenStreetMap

OpenStreetMap



Ordnance Survey



Networks

Detailed and accurate network datasets for highways, roads, paths, and waterways.

[Explore networks](#)

Topography

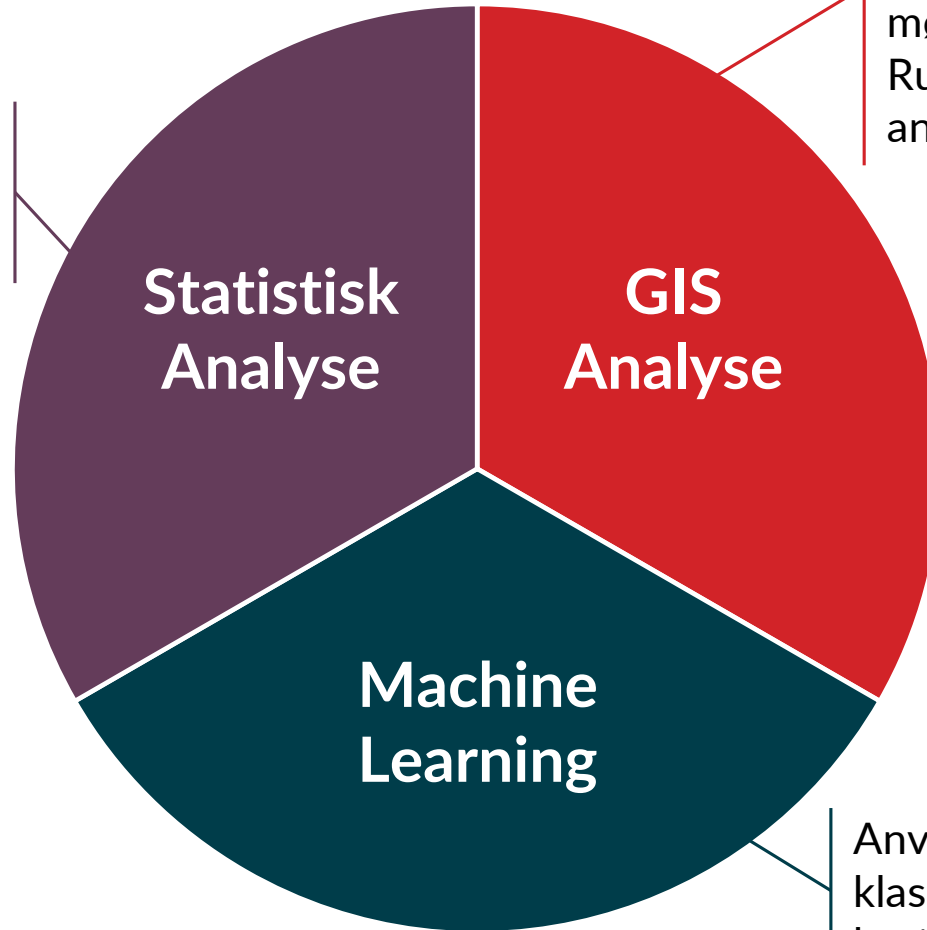
Detailed topographical products help visualise geographic data to make clearer location-based decisions about your assets, risks and services.

[Explore topography](#)

SYSTRA

Metoder

Kategorisering vha. Cluster Analyse
Binær logistisk regression
PCA

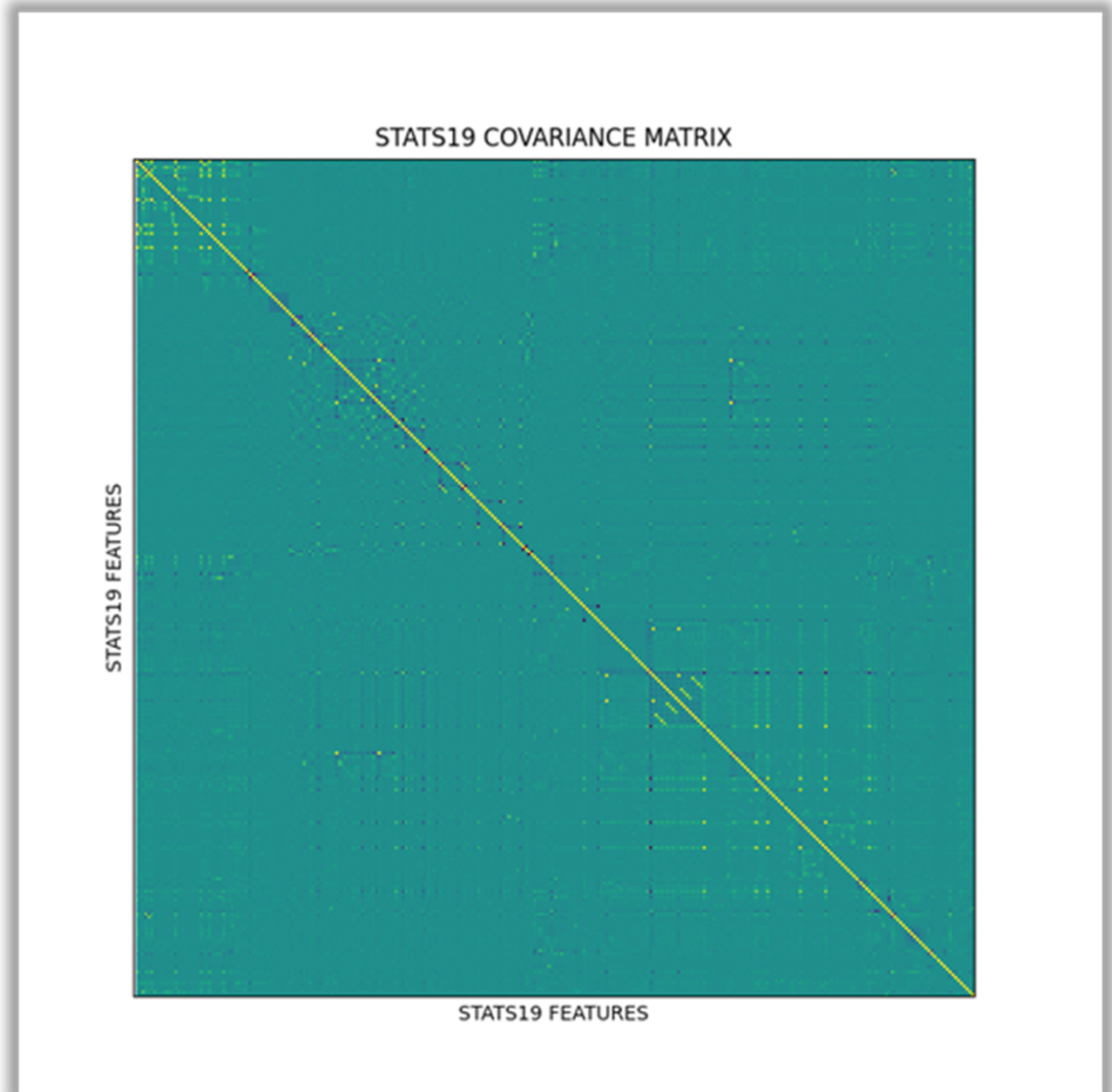


Hvor? og Hvorfor?
Kortlægning af geografiske mønstre i alvorlige kollisioner.
Rumlige autokorrelation, hot spot analyse, tendensanalyse.

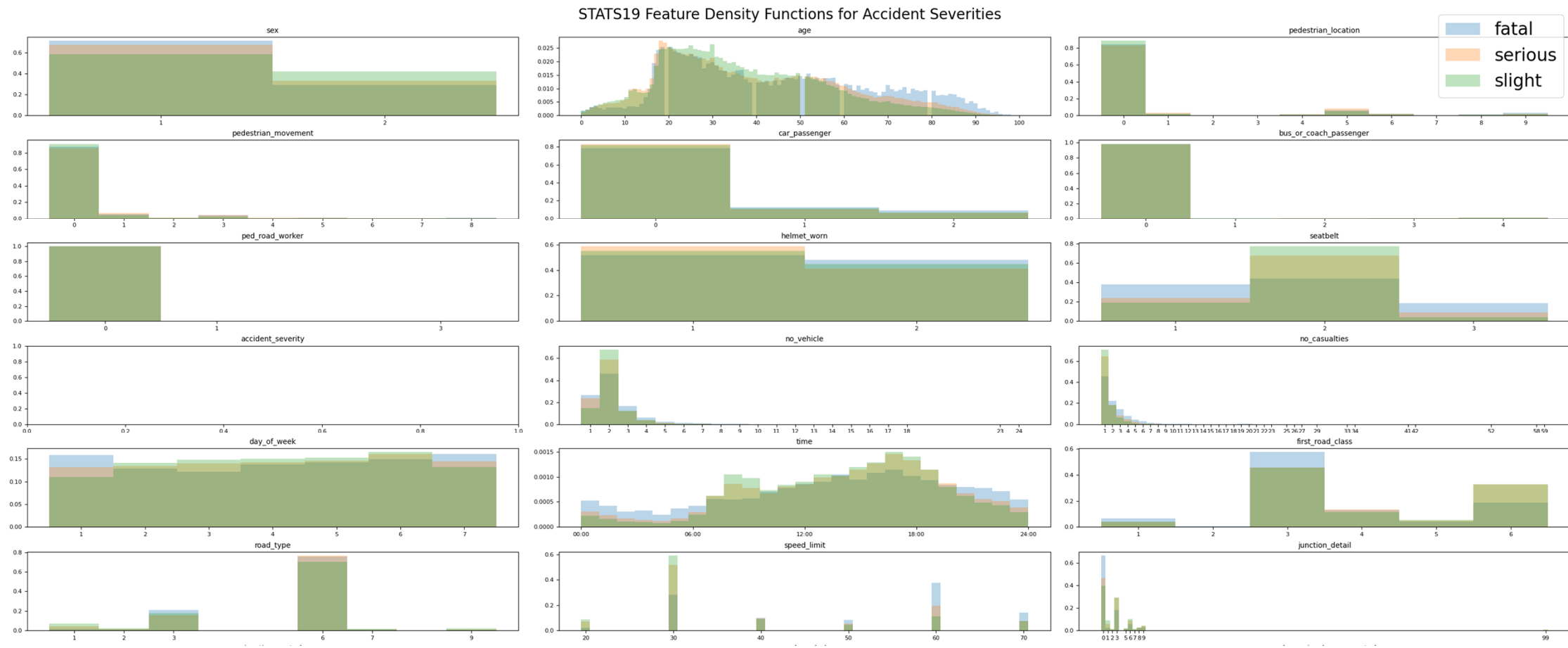
Anvendte ML teknikker til klassificering af det fysisk kortlagte miljø som defineret i OSM og OS MasterMap

Statistisk Analyse

- Holistisk Analyse
- Gruppering Cluster Algoritmer
- Køretøjstyper
- Temporal Variation

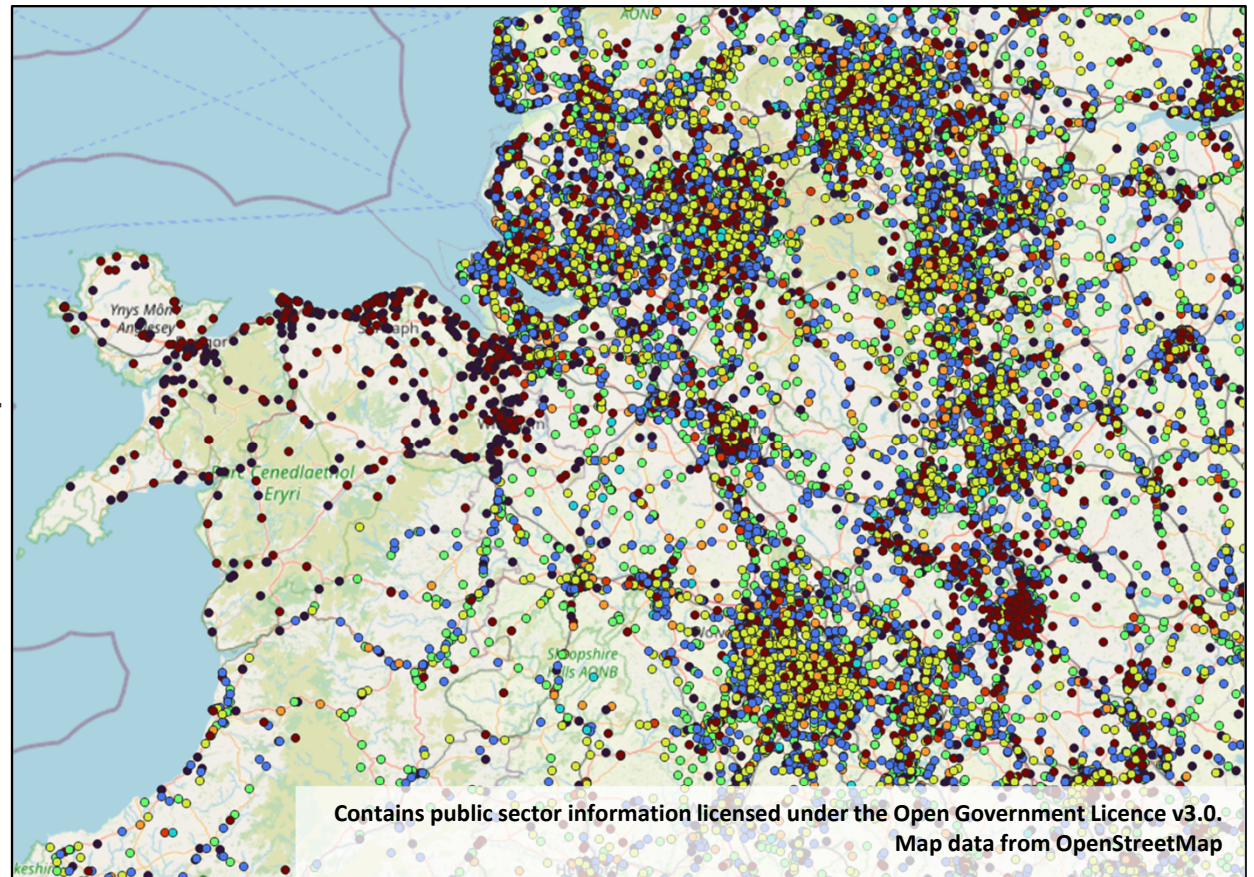


Detaljeret overblik

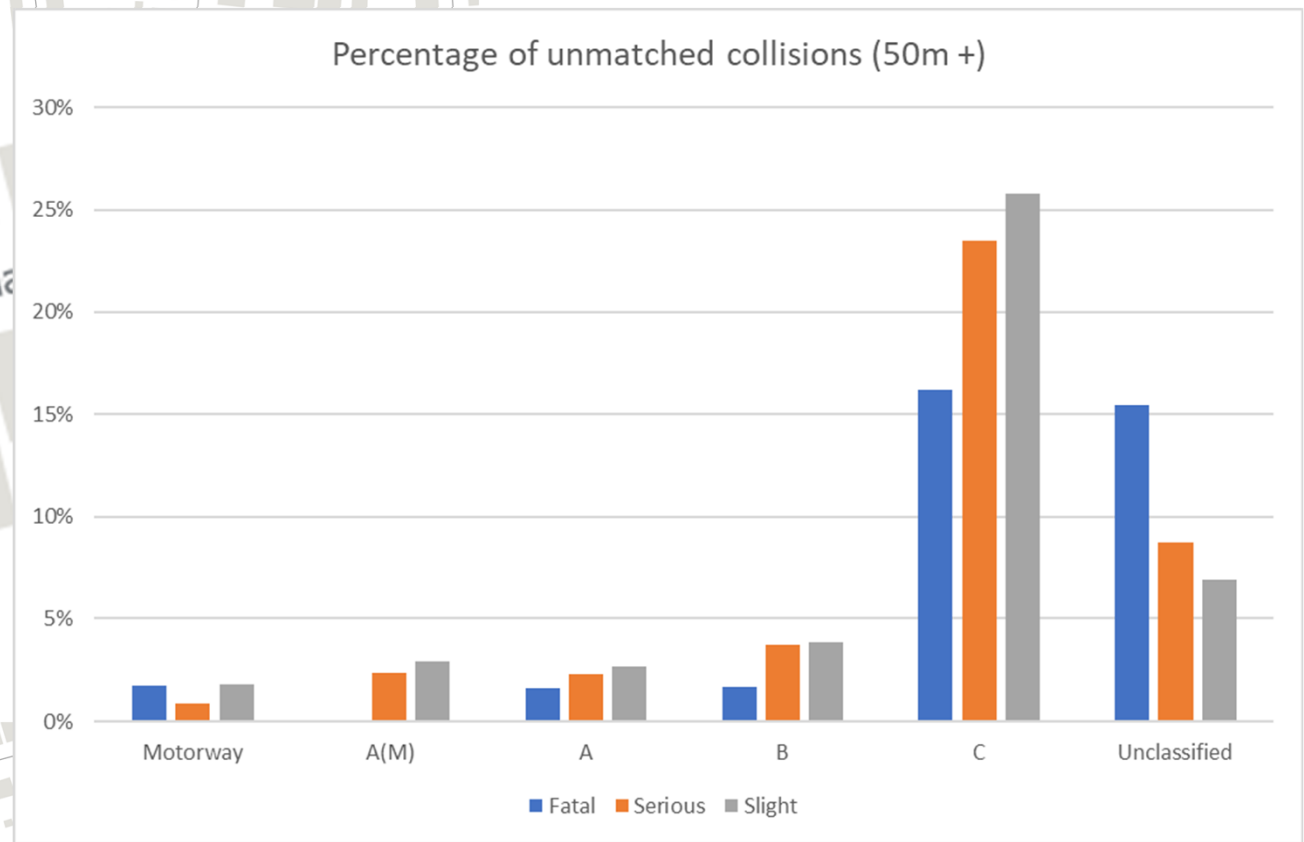
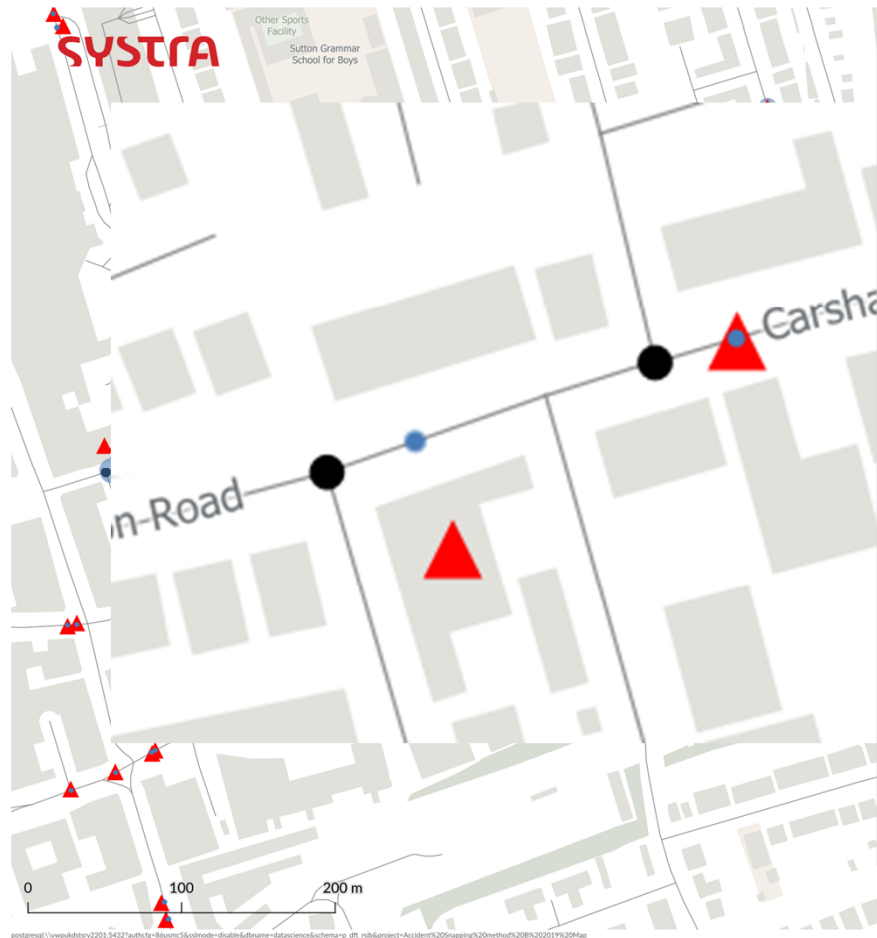


Gruppering som data check

- 32 variable i standard data
- 10 sensitive variable
- Binær repræsentation vel over 500 variable
- Første iterationer af kluster analyser viste skjulte geografiske data karakteristika
 - Ledte til data reduktion
 - Indfyldning etc.



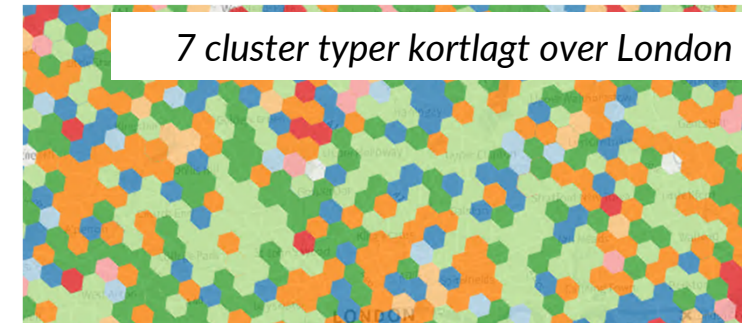
Forbedret placering af kollisioner - smart snapping



Contains OS data © Crown copyright and database right 2024
Contains public sector information licensed under the Open Government Licence v3.0

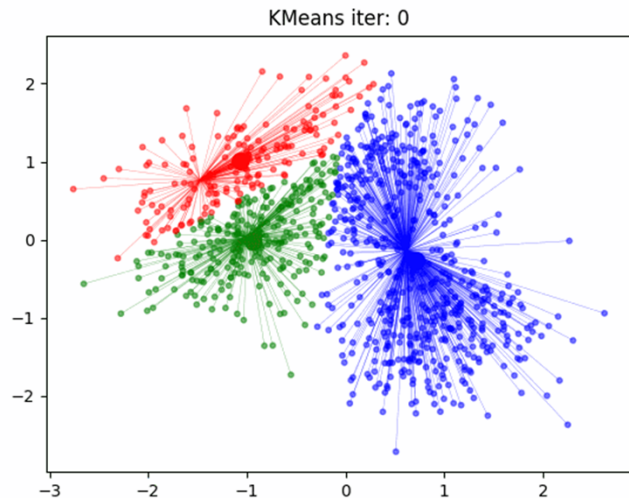
Gruppering – K Mode Clustering og Binær Regression

- Vejkollisioner er (efterhånden) tilfældige hændelser (resulterer i data støj).
 - Traditionel sort plet analyse er færdiggjort.
 - Vi har forklaret og udbedret hvad vi umiddelbart kan forstå og eftervise som funktion af vejdesign.
- K-modes analyse afslører nye komplekse mønstre.
 - 7 klynger opstod i første iteration.
- Regressionsanalyse for hver klynge giver
 - Forståelse for hvilke faktorer der best forklarer skadesgraden ved en kollision i gruppen



K-Modes Cluster algoritme

- Unsupervised machine learning metode til at gruppere dataset



4 grupper

FLERE INVOLVEREDE KØRETØJER

- Dødelige udgang, hvor køretøjer vælter
- Højere risiko for dødelige udgang i rundkørsler hvor godskøretøjer er involveret
- Alvorlige hændelser, der involverer flere køretøjer

BERUSEDE FØRERE

- Cykelister, der er berusede og ikke bærer hjelm, er i betydelig risiko for dødelig udfald.
- Berusede førere, der er involveret i dødelige hændelser, kolliderer med genstande på kørebanen, vejarbejder, tidligere kollisioner og bilister der skrider ud og vælter

CHAUFFØRER, UDEN SPÆNDT SIKKERHEDSSELE

- Hændelser, der involverer ældre køretøjer, er mere dødelige, måske pga kvaliteten eller holdbarheden af sikkerhedsfunktioner i biler (som airbags)
- Dødelige og alvorlige hændelser, hvor køretøjer er væltet

ÆLDRE BILISTER / TILSKADEKOMNE

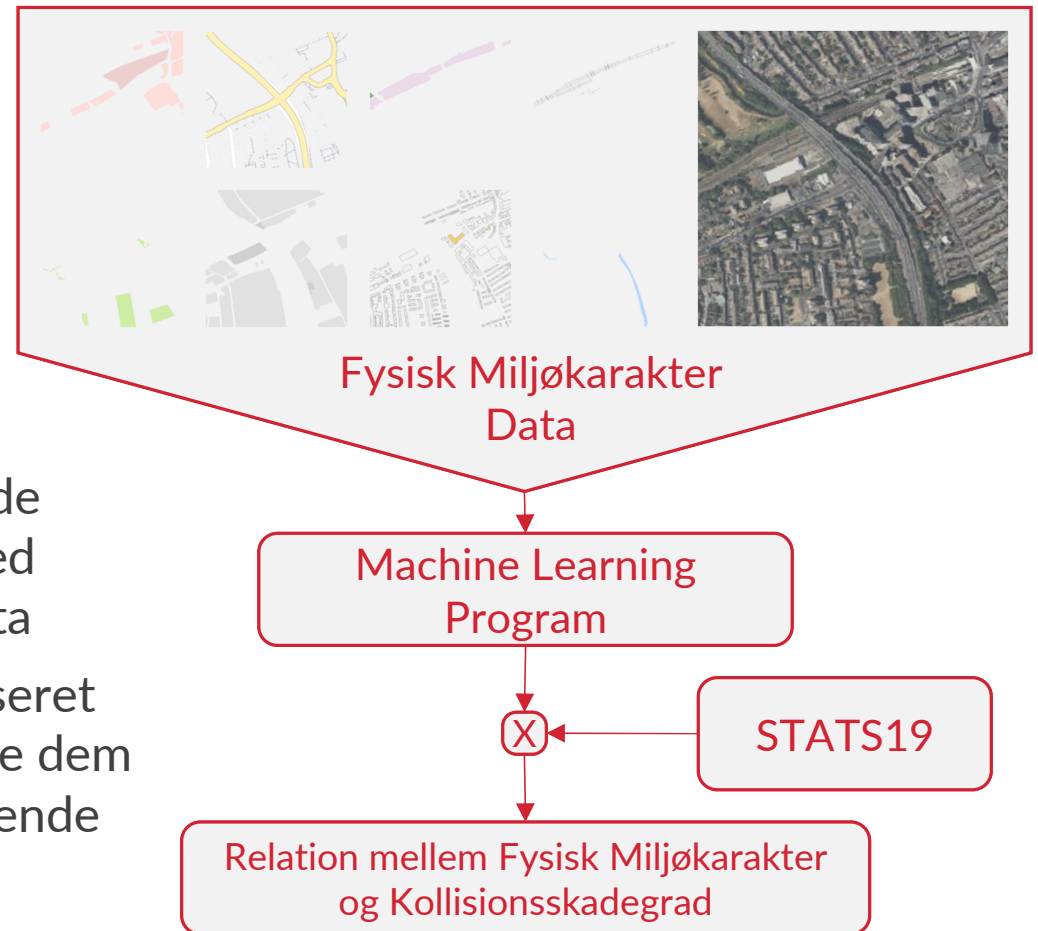
- Hændelser, der involverer ældre cykelister, er mere tilbøjelige til at være fatale eller alvorlige, især for dem, der ikke bærer hjelm
- Ældre bilister, uden spændt sikkerhedssele

Bidragende Faktorer for Køretøjstyper efter størst forklarende værdi

Rank	Car	Motorcycle over 500cc	Bus	Pedal Cycle	Tram
1	Rural Road	60mph Speed Limit	Rural Road	Rural Road	Dark Conditions with No Lighting
2	60mph Speed Limit	Front on Collision	60mph Speed Limit	60mph Speed Limit	Unknown Road Type
3	Dark Unlit Road	Rural Road	Front on Collision	Not Wearing a Helmet	Road With Pedestrian Central Refuge
4	Not at or adjacent to Junction	Unknown Purpose of Journey	Age of Driver (Older)	Not at, or in vicinity of, a Junction	Pedestrian Hit Within Carriageway, Not Crossing
5	Seatbelt Worn	Not at, or in vicinity of, road junction	Dark Unlit Road	Cyclist Hit at Rear	Male Sex
6	Leaving Carriageway on Nearside	Going Ahead Left-Hand Bend	Leaving Carriageway on Nearside	Leaving Carriageway on Nearside	Roads Speed Limit of 20mph
7	Collision with Parked Vehicle	Collisions Involving 3 Vehicles	Overtaken Vehicle	Dark Unlit Road	Tram On Right Hand Bend
8	Leaving Carriageway on Offside	Leaving Carriageway on Nearside	Wet or Damp Road Surface	Vehicle Overtaken	Front on Collision
9	Going Ahead Right-Hand Bend	Collisions Involving 4 Vehicles	Leaving Carriageway on Offside	Collisions Involving 3 Vehicles	Dual Carriageway
10	Going Ahead Left-Hand Bend	Single Carriageway Road	Vehicle Age Between 20 and 25 Years Old	50mph Speed Limit	Failed Breath Test

Machine Learning: Overblik

- Hypotese at det **fysiske miljø** har **forklarende værdi** i de medvirkende faktorer i kollisionshændelser – men er ikke velbeskrevet i STATS19 data.
- **Semantisk vektorindlejring** udført for at **identificere og kategorisere** kollisionsteders fysiske miljø
- Implementation of 2 **neurale netværk** til at finde semantiske ligheder mellem kollisionsteder ved brug af OpenStreetMap og OS MasterMap data
- Som resultat kan vi nu gruppere kollisioner baseret på **ligheder i deres fysiske miljøer** og analysere dem rumligt for at evaluere miljøet som medforklarende variabel for alvorlige kollisionsskade udfald.

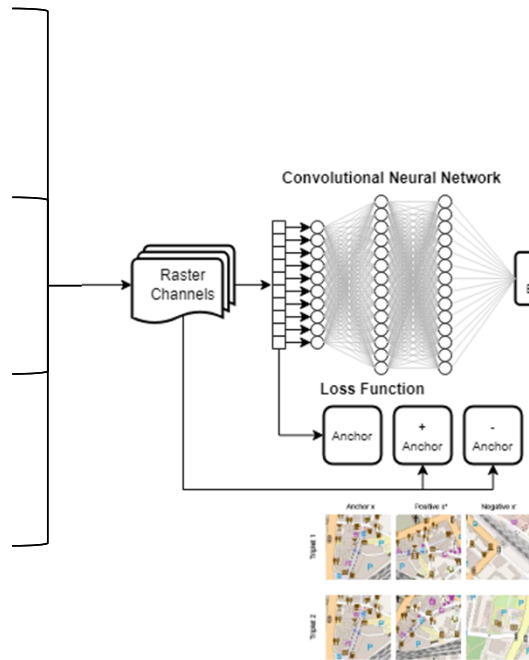


Machine Learning Model Opbygning

KOLLISIONS
STED



ML
MODEL



VEKTOR
REPRESENTATION

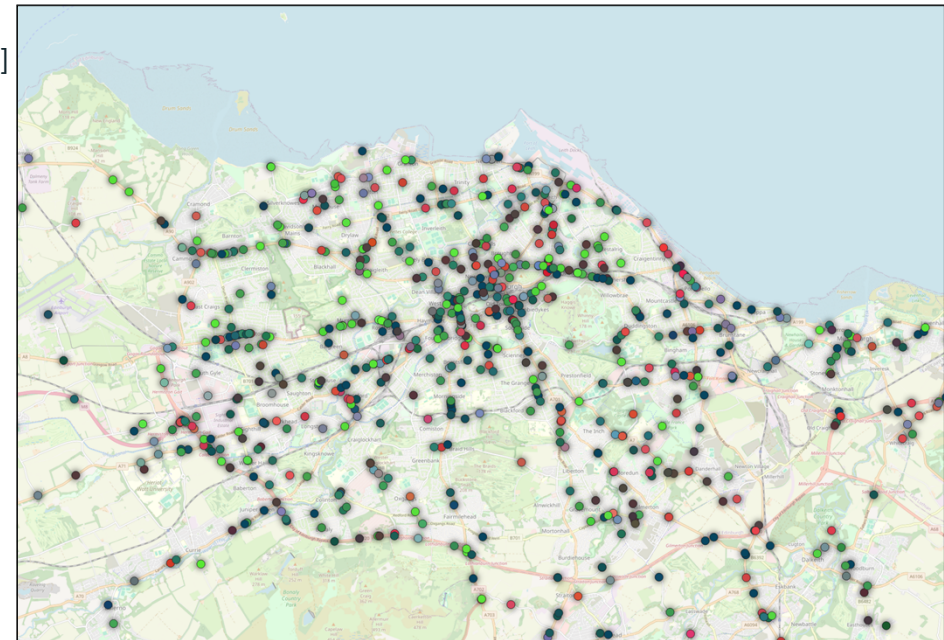
[0.3214, 3.123, ..., 0.834]

[6.914, 1.203, ..., 4.231]

[1.114, 7.223, ..., 0.189]

[2.011, 2.121, ..., 0.072]

N-DIMENSIONAL REGRESSION KORTLAGT
PCA KONVERTERET TIL RGB



Machine Learning: Træning

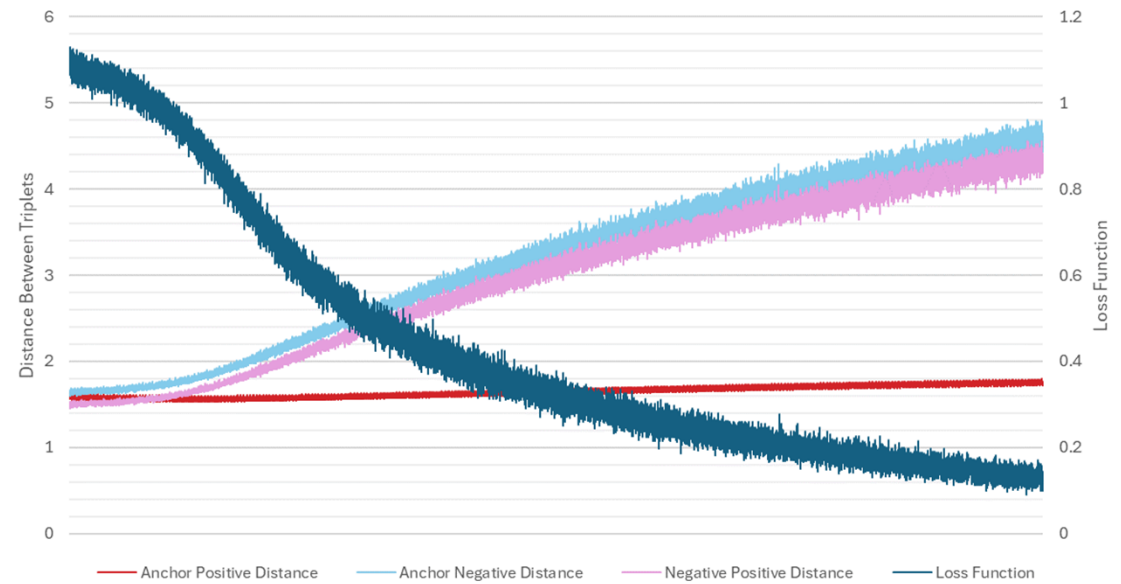
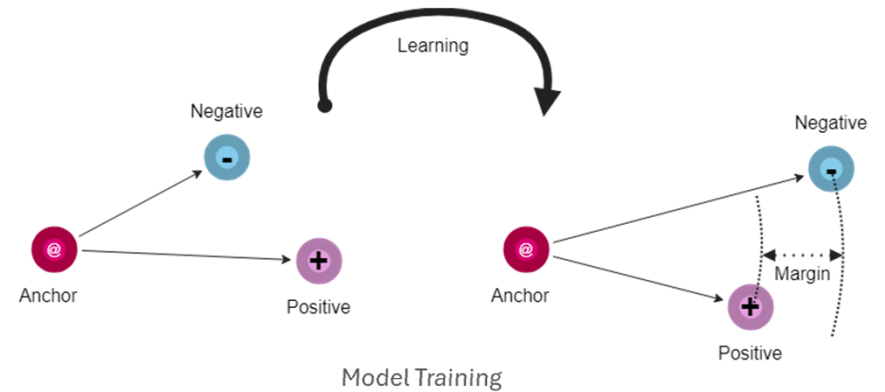
ANCHOR



ANCHOR | POSITIVE



ANCHOR | NEGATIVE



Machine learning: Resultater

3 af 7 kategorier af kollisioner
baseret på det fysiske miljø.
(5 bymæssige og 2 landdistrikter)



Opsummering

1. Hændelser er generelt koncentreret i og omkring tættere befolkede områder, mens dødelige og alvorlige hændelser er stærkt koncentreret på landeveje, især hvor de er ubelyste og underlagt den nationale hastighedsgrænse.
2. Skarpe sving på landeveje og køretøjer, der foretager svingmanøvrer i høj fart ind på en anden vej, er nøgleomstændigheder, hvor dødelige og alvorlige hændelser opstår.
3. Ved at opdele dataene i klynger, køretøjstyper og fysiske miljøer opnås yderligere indsigt i omstændigheder og medvirkende faktorer, der kan overses, når man betragter det bredere datasæt.
4. Klyngeanalyse skaber 'kategorier' af kollisioner, som kan hjælpe med at identificere foranstaltninger der kan mindske alvorligheden af fremtidige kollisioner.
5. Geospatial analyse og visualiseringer identificerer lokationer, der bør prioriteres ved implementering af afbødningsforanstaltninger.
6. Maskinlæringsteknikker har identificeret, at steder, hvor strategiske veje skærer gennem forstadsbeboelsesområder med begrænsede formelle krydsningsmuligheder, øger sandsynligheden for dødelige og alvorlige skader.

SYSTRA

Yderligere information:

systra.com/uk

systra.com/ireland

Systra.com/denmark



CONFIDENCE MOVES THE WORLD