



Dedicated to innovation in aerospace

NLR-TR-2020-379 | maart 2021

Data voor de Nationale Veiligheidsanalyse (NVA)

Onderzoek naar databronnen, dataverrijking en Machine Learning
toepassingen voor de NVA

Kennis voor Beleid



NLR – Koninklijk Nederlands Lucht- en Ruimtevaartcentrum

Data voor de Nationale Veiligheidsanalyse (NVA)

Onderzoek naar databronnen, dataverrijking en Machine Learning toepassingen voor de NVA



Probleemstelling

Het Ministerie van Infrastructuur en Waterstaat stelt met behulp van de periodieke nationale veiligheidsanalyse (NVA) de potentiële gevaren en risico's van nationaal belang vast, en ontwikkelt maatregelen om deze risico's te beheersen. De NVA omvat de identificatie en analyse van nationale veiligheidsrisico's voor de commerciële, kleine en onbemande luchtvaart. Om deze analyse te kunnen uitvoeren, is objectieve informatie nodig op basis van data. In het kader van het Kennis voor Beleid programma (de vraaggestuurde kennisontwikkeling voor de overheid) heeft het NLR een onderzoek uitgevoerd naar het gebruik van veiligheids- en operationele data ten behoeve van een NVA.

Beschrijving van de werkzaamheden

Het doel van het onderzoek was het identificeren en beoordelen van databronnen die gebruikt kunnen worden voor gevaarsidentificatie, risicobeoordeling en risicomitigatie in een NVA. Tijdens de analyse van data kan het zinvol zijn aanvullende informatie of gegevens te gebruiken om de data te verrijken of te normaliseren om de context in de analyse mee te kunnen nemen. In dit onderzoek

RAPPORTNUMMER

NLR-TR-2020-379

AUTEUR(S)

R. Wever
V.J.F. de Vries
P.B.M. Jonk
R. Kramer
J. Choi
R.P.M. Verhoeven

RUBRICERING RAPPORT

ONGERUBRICEERD

DATUM

maart 2021

KENNISGEBIED(EN)

Luchtvaartveiligheid

TREFWOORD(EN)

Dataverrijking
Veiligheidsanalyse
Machine Learning

is daarom een lijst met mogelijke contextuele data voor verrijking of normalisatie opgesteld op basis van beschikbare kennis, ervaring en eerdere veiligheidsanalyses. Daarnaast zijn de mogelijkheden verkend van de toepassing van Machine Learning (ML) technieken om data-analyse in de NVA te ondersteunen. Naast een literatuurstudie naar de toepassing van ML heeft het NLR diverse ML-technieken uitgetest in drie test cases, namelijk: 1) het categoriseren van incidentenrapporten met een 'occurrence category', 2) het samenvatten van de 'probable cause' van een voorval op basis van teksten uit het onderzoeksrapport van het voorval, en 3) een sentimentanalyse op Twitter berichten over luchtvaart.

Resultaten en conclusies

Er zijn 57 databronnen geïdentificeerd die geraadpleegd kunnen worden in de NVA. Deze bronnen beschrijven verschillende elementen van het Nederlandse luchtvaartstelsel, en beslaan de commerciële, kleine en onbemande luchtvaart. Aan de hand van een beoordelingskader zijn de databronnen gecategoriseerd op verschillende karakteristieken die een eerste indruk geven van de compleetheid, toegankelijkheid, dekkingsgraad en reikwijdte van de databronnen. In deze studie zijn tevens een aantal criteria gedefinieerd die als leidraad kunnen dienen voor de analist om de kwaliteit van de data te beoordelen. De lijst met contextuele data geeft de mogelijkheden weer voor verrijking en normalisatie van veiligheidsdata. Er zijn enkele voorbeelden gegeven van het verrijken van data ter illustratie van de meerwaarde. Op basis van de toepassing van ML-technieken in drie use cases en de eerste resultaten wordt geconcludeerd dat deze technieken goede mogelijkheden bieden voor ondersteuning van veiligheidsanalyses. De ML-technieken lenen zich vooral voor het analyseren van grote datasets en het voorspellen van bepaalde uitkomsten op basis van historische data. Het toepassen van ML vraagt naast kennis van 'data science' ook domeinkennis van luchtvaart(veiligheid) om de ML-modellen te ontwikkelen, te trainen en te testen, en vervolgens de resultaten op waarde te kunnen beoordelen.

Toepasbaarheid

De geïdentificeerde databronnen en contextuele data kunnen informatie leveren voor de analyse in de NVA. Er zijn in dit onderzoek een aantal criteria geformuleerd, als leidraad voor de analist, om de kwaliteit van de data(bron) voor het betreffende criterium te beoordelen. Daarnaast zijn diverse ML-technieken bruikbaar ter ondersteuning van de analyse, bijvoorbeeld voor het categoriseren van voorvallen of het analyseren van grote hoeveelheden voorvalrapporten.

NLR

Anthony Fokkerweg 2

1059 CM Amsterdam

p) +31 88 511 3113

e) info@nlr.nl i) www.nlr.nl



Data voor de Nationale Veiligheidsanalyse (NVA)

Onderzoek naar databronnen, dataverrijking en Machine Learning toepassingen voor de NVA

Kennis voor Beleid

AUTEUR(S):

R. Weber	NLR
V.J.F. de Vries	NLR
P.B.M. Jonk	NLR
R. Kramer	NLR
J. Choi	NLR
R.P.M. Verhoeven	NLR

Niets uit dit rapport mag worden vermenigvuldigd en/of openbaar gemaakt, op welke wijze dan ook, zonder voorafgaande schriftelijke toestemming van het NLR.

OPDRACHTGEVER	-----
CONTRACTNUMMER	-----
EIGENAAR	NLR
NLR DIVISIE	Aerospace Operations
VERSPREIDING	Onbeperkt
RUBRICERING TITEL	ONGERUBRICEERD

GOEDGEKEURD DOOR:		Datum
AUTEUR	R. Wever	02-03-2021
REVIEWER	S.H. Stroeve	02-03-2021
BEHERENDE AFDELING	A.D.J. Rutten	02-03-2021

Samenvatting

Het Ministerie van Infrastructuur en Waterstaat stelt met behulp van een periodieke nationale veiligheidsanalyse (NVA) de potentiële gevaren en risico's van nationaal belang vast, en ontwikkelt maatregelen om deze risico's te beheersen. De NVA omvat de identificatie en analyse van nationale veiligheidsrisico's voor de commerciële, kleine en onbemande luchtvaart. Om deze analyse te kunnen uitvoeren, is objectieve informatie nodig op basis van data. In het kader van het Kennis voor Beleid programma (de vraaggestuurde kennisontwikkeling voor de overheid) heeft het NLR een onderzoek uitgevoerd naar het gebruik van veiligheids- en operationele data ten behoeve van de NVA. Het doel van het onderzoek was het identificeren en het beoordelen van databronnen die gebruikt kunnen worden voor gevaarsidentificatie, risicobeoordeling en risicomitigatie in de nationale veiligheidsanalyse. Daarnaast was het doel om in kaart te brengen welke verrijkingen en normalisatie van data mogelijk zijn om context te bieden bij de data tijdens een analyse. Tenslotte zijn de mogelijkheden onderzocht van een aantal Machine Learning (ML) technieken om de data-analyse in de NVA te ondersteunen.

Identificatie van databronnen

De data(bronnen) moeten helpen de volgende vragen te beantwoorden in de NVA:

- Wat zijn de gevaren in de Nederlandse luchtvaart voor de commerciële, kleine en onbemande luchtvaart?
- Wat zijn de grootste risico's voor deze drie domeinen?
- Hoe komt het dat deze risico's in het luchtvaartstelsel bestaan? Welke factoren spelen een rol?
- Welke maatregelen zijn nodig om de risico's, gegeven de context waarin ze optreden, te beheersen?

Er is een aanzienlijke hoeveelheid databronnen beschikbaar voor de NVA. Deze studie heeft 57 databronnen geïdentificeerd die geraadpleegd kunnen worden voor gevaarsidentificatie, risicobeoordeling en risicomitigatie. Deze bronnen beschrijven verschillende elementen van het Nederlandse luchtvaartstelsel en beslaan de domeinen commerciële, kleine en onbemande luchtvaart. Aan de hand van een beoordelingskader zijn de databronnen gecategoriseerd op verschillende onderwerpen. Dit maakt belangrijke karakteristieken van de databronnen inzichtelijk en geeft een eerste indruk van de compleetheid, toegankelijkheid, dekkingsgraad en reikwijdte van de databronnen. De dekkingsgraad van databronnen voor de domeinen kleine luchtvaart (General Aviation) en onbemande luchtvaart (drone operations) is een stuk lager dan voor de commerciële luchtvaart. De databronnen dekken alle elementen in het luchtvaartstelselmodel af. Het merendeel van de bronnen geeft informatie over de (inrichting van de) vliegoperatie en de prestatie van de operatie. Circa één derde van de databronnen biedt informatie over de het toezicht op de operatie en het lerend vermogen van het luchtvaartstelsel.

De sectorpartijen beschikken over databronnen die ruwe (niet verwerkte of geanalyseerde) data bevatten, zoals vluchtdata, radardata, en voorvalmeldingen. Deze data zijn zeer nuttig ter ondersteuning van een NVA. Het vraagt echter een aanzienlijke inspanning en middelen om deze data te verzamelen, te verwerken, en te analyseren voor een risicoanalyse. Daarom wordt aanbevolen om in de NVA-methodiek richtlijnen op te stellen voor welke soort vraagstukken het de moeite waard is om dergelijke ruwe data te benutten.

De overheid beschikt met name over databronnen die statistiek en analyses bevatten. Deze informatie is eenvoudiger toegankelijk en laagdrempeliger te gebruiken in een NVA dan ruwe data. Echter, het gebruik van bronnen met statistiek en analyses moet zorgvuldig gebeuren omdat in de onderliggende bronnen aannames, keuzes en interpretaties zijn gemaakt die invloed hebben op de toepasbaarheid of representativiteit van de bronnen voor de NVA. Bovendien zijn gegevens uit verschillende statistieken meestal lastig te combineren vanwege de verschillende context en dimensies.

De kwaliteit en betrouwbaarheid van de data en informatie in de bronnen varieert sterk. De kwaliteit van data wordt bepaald door de mate waarin data actueel, nauwkeurig, betrouwbaar, representatief, en volledig is, en daarnaast voldoende informatie en detailniveau bevat om een kwalitatief goede analyse uit te kunnen voeren. Er zijn in deze studie een aantal criteria geformuleerd, als leidraad voor de analist, om de kwaliteit van de data(bron) voor het betreffende criterium te beoordelen.

Verrijking en normalisatie van data

Tijdens de analyse van data kan het zinvol zijn aanvullende informatie of gegevens te gebruiken om de data te verrijken of te normaliseren, om zo de context in de analyse mee te nemen. Verrijking betekent dat een databron door het toevoegen van andere data(bronnen) 'rijker' aan informatie wordt. Een voorbeeld is het toevoegen van de weerscondities en landingsbaan aan gemelde go-arounds. Normaliseren houdt in dat het aantal observaties, bijvoorbeeld het aantal incidenten, afgezet wordt tegen een bepaalde context, bijvoorbeeld het weer of een type operatie. Bijvoorbeeld, het aantal incidenten per 10.000 vluchten, per maand, of in een bepaalde zichtconditie.

Verrijking en normaliseren helpen inzichtelijk te maken hoe groot risico's zijn, welke factoren een rol spelen en hoe groot die rol is, en welke relaties er (mogelijk) bestaan tussen factoren. De dataverrijking en normalisatie die mogelijk en zinvol is, is sterk afhankelijk van het vraagstuk, de gewenste analyse en de beschikbare data. De kennis en ervaring van de analist speelt ook een belangrijke rol in de keuze van de specifieke verrijking en normalisatie.

In dit onderzoek heeft het NLR een conceptlijst van mogelijke contextuele data voor verrijking en normalisaties opgesteld op basis van beschikbare kennis, ervaring en eerdere veiligheidsanalyses. Deze contextuele data bieden informatie over de mate van activiteit of productie van een partij, en/of informatie over het 'systeem' en de omgeving waarin de luchtvaartoperatie plaatsvindt. In het algemeen kan men stellen dat normalisatie naar activiteit (aantal vluchten, aantal landingen, aantal inspecties, etc.), naar tijdseenheid (per jaar, kwartaal, maand, etc.) of omgevingsfactoren (weerscondities, locatie, vluchtfase, baangebruik) een goede eerste stap is. Enkele voorbeelden van verrijking en normalisatie die in dit rapport zijn beschreven, laten zien hoe divers verrijking en normalisatie kunnen zijn en welke aanvullende inzichten het kan opleveren.

Toepassing van Machine Learning

Een deel van dit onderzoek richtte zich op de toepassing van Machine Learning (ML) ten behoeve van de NVA. Op basis van een literatuurstudie zijn een aantal interessante toepassingen en analyses in kaart gebracht, die mogelijk op termijn ingezet kunnen worden om de analyse in de NVA te ondersteunen. De verkenning van Machine Learning technieken omvatte ontwikkelingen bij andere organisaties in de luchtvaart, maar ook toepassingen buiten de luchtvaart die mogelijk bruikbaar zijn voor de NVA. In de literatuurstudie naar Machine Learning zijn zes relevante toepassingen gevonden. Deze zijn kort beschreven en geclassificeerd naar onderwerp, categorie en methode van Machine Learning, input en output, en mogelijke toepassing en relevantie voor de NVA.

Gelet op de reikwijdte van ML-technieken en mogelijke toepassingen is in dit onderzoek gekozen voor de toepassing van text-mining technieken ter ondersteuning van de NVA. Het NLR-team heeft enkele Machine Learning technieken op het gebied van text-mining toegepast op drie use cases om de haalbaarheid en bruikbaarheid van deze technieken te evalueren. De cases betroffen 1) het categoriseren van voorvalmeldingen met een 'occurrence category', 2) het samenvatten van de 'probable cause' van een voorval op basis van teksten uit het onderzoeksrapport van het voorval, en 3) een sentimentanalyse op Twitter berichten over luchtvaart. De tweede case is in feite een verwerkingsstap ter voorbereiding van een analyse van oorzaken en verbanden tussen oorzakelijke factoren in een grote verzameling van ongevallen- en incidentenrapporten.

De toepassing van de ML-technieken heeft veelbelovende resultaten opgeleverd. Op basis van de toepassing van ML-technieken in drie use cases zijn per case een aantal conclusies en aanbevelingen geformuleerd. Daarnaast worden ook enkele algemene conclusies getrokken over de toepassing van Machine Learning:

- Op basis van de drie cases studies en de resultaten van de literatuurstudie lijken ML-technieken goede mogelijkheden te bieden voor ondersteuning van veiligheidsanalyses. De technieken lenen zich vooral voor het analyseren van grote datasets en het voorspellen van bepaalde uitkomsten op basis van historische data.
- Het toepassen van ML vraagt naast kennis van 'data science' ook domeinkennis van luchtvaart(veiligheid) om zowel de ML-modellen te ontwikkelen, te trainen en te testen, maar ook om de resultaten op waarde te kunnen beoordelen.
- Het finetunen van de modellen en het maximaliseren van de opbrengst van de ML-technieken kost veel tijd. De pre-processing en finetuning spelen een belangrijke rol in het analyseproces en leggen een aanzienlijk beslag op de capaciteit. Het ligt daarom voor de hand om ML vooral in te zetten daar waar het de efficiëntie kan verhogen, bijvoorbeeld voor het vervangen van regulier handmatig classificeren van data.
- De ontwikkeling en training van een model wordt beter en gaat sneller als de analist beschikt over een gelabelde dataset. De bruikbaarheid van trainingsdata van andere bronnen of uit andere toepassingen was beperkt in dit onderzoek door afwijkingen in datastructuur, type, inhoud en/of format. De analist kan in deze situatie kiezen om een eigen trainingset ontwikkelen of een handmatige validatie achteraf uit voeren op de resultaten. Het gebrek aan gelabelde data in dit onderzoek betekent dat de validatie van de resultaten op dit moment handwerk is en veel inspanning vraagt.

Het NLR heeft in dit onderzoek een aantal aanbevelingen opgesteld voor de verdere ontwikkeling van de toepassing van ML-technieken in de drie cases. De gebruikte ML-modellen kunnen verder ontwikkeld worden om de prestatie in de betreffende cases te verbeteren waardoor de kwaliteit van de resultaten zal verbeteren. Een belangrijke stap in deze ontwikkeling is de validatie van de resultaten. Daarnaast wordt aanbevolen om het onderzoek naar de toepassing van ML ook te richten op het operationaliseren van de toegepaste technieken. Het regulier toepassen van deze ML-technieken door het NLR of andere partijen (denk aan bijvoorbeeld het ABL) vraagt verder onderzoek naar o.a. de praktische implementatie en instandhouding. Tenslotte raden we aan om verder onderzoek te doen naar andere ML-technieken voor de analyse van andere soorten data. Met name de toepassing van ML-technieken op radar data of vluchtdata is interessant met het oog op bijvoorbeeld het detecteren van afwijkende vluchten, het duiden van trends, en het monitoren van de effecten van genomen risicobeheersmaatregelen.

Inhoudsopgave

Afkortingen	8
1 Introductie	9
1.1 Achtergrond	9
1.2 Doel van het onderzoek	9
1.3 Scope van dit onderzoek	9
1.4 Het Nederlandse luchtvaartstelsel	10
1.5 Leeswijzer	11
2 Databronnen	12
2.1 Inleiding	12
2.2 Aanpak	12
2.3 Resultaten	14
2.3.1 Lijst met databronnen	14
2.3.2 Kwaliteit	14
2.3.3 Dekkingsgraad	14
2.3.4 Eigendom en toegankelijkheid	17
2.4 Kwaliteit van data en waarde voor de NVA	18
2.4.1 Achtergrond	18
2.4.2 Kwaliteit van databronnen	18
2.4.3 Waarde van databronnen voor de NVA	19
2.4.4 Combineren van databronnen voor de NVA	21
2.5 Beschouwing	22
3 Dataverrijking en normalisering	23
3.1 Inleiding	23
3.2 Resultaten	23
3.2.1 Overzicht van contextuele data	23
3.2.2 Voorbeelden van verrijking en normalisatie	24
3.3 Beschouwing	29
4 Machine Learning	30
4.1 Inleiding	30
4.2 Resultaten literatuurstudie	30
4.3 Toepassing van ML in drie use cases	31
4.4 Case 1: Automatisch labelen van voorvalrapport met de Occurrence Category	32
4.4.1 Inleiding	32
4.4.2 Data	33
4.4.3 Preprocessing	33
4.4.4 Model	34
4.4.5 Resultaat	34
4.4.6 Conclusie	36
4.4.7 Aanbevelingen	36

4.5	Case 2: Automatisch genereren van een probable cause op basis van een feitenbeschrijving	36
4.5.1	Inleiding	36
4.5.2	Data	36
4.5.3	Tokenizer	37
4.5.4	Pre-processing	38
4.5.5	Model	39
4.5.6	Resultaten	41
4.5.7	Conclusie	45
4.5.8	Aanbevelingen	45
4.6	Case 3: Sentiment analyse op sociale media berichten	46
4.6.1	Inleiding	46
4.6.2	Definitie van sentiment analyse	46
4.6.3	Data	47
4.6.4	Model 1: Supervised sentiment analyse	47
4.6.5	Model 2: Unsupervised sentiment analyse	48
4.6.6	Resultaten	50
4.6.7	Conclusies	53
4.6.8	Aanbevelingen	54
5	Conclusies en aanbevelingen	55
5.1	Conclusies	55
5.2	Aanbevelingen	56
6	Referenties	57
Appendix A	Overzicht van databronnen met categorisering	60
Appendix B	Lijst met voorbeelden van contextuele data	72
Appendix C	Literatuuronderzoek toepassing Machine Learning	75
Appendix C.1	Introduction	75
Appendix C.2	Introduction to Machine Learning	75
Appendix C.3	Types of Machine Learning algorithms	76
Appendix C.4	Data for ML application	77
Appendix C.5	Machine Learning Use Cases	78
Appendix C.5.1	Case 1: Comparison of influencing factors for the flight efficiency	78
Appendix C.5.2	Case 2: Prediction of pilot misperception on runway excursion risk	79
Appendix C.5.3	Case 3: Understanding accident characteristics and contributing factors	80
Appendix C.5.4	Case 4: Text classification of occurrence reports	80
Appendix C.5.5	Case 5: Public opinion analysis using classification	81
Appendix C.5.6	Case 6: Detection of anomalies in flight operations	82
Appendix D	Zoektermen Twitter API	83

Afkortingen

ACRONIEM	OMSCHRIJVING
ABL	Analysebureau luchtvaartvoorvallen
ANSP	Air Navigation Service Provider
ARBO	Arbeidsomstandigheden
BES	Bonaire Sint Eustatius Saba
CAT	Commercial Air Transport
CICTT	CAST/ICAO Common Taxonomy Team
EASA	European Union Aviation Safety Agency
GA	General Aviation
ICAO	International Civil Aviation Organisation
IenW	Ministerie van Infrastructuur en Waterstaat
ILT	Inspectie Leefomgeving en Transport
ISMS	Integraal Safety Management Systeem (joint sector SMS voor Schiphol)
ISMS TOPSAG	ISMS Top Safety Action Group
ML	Machine Learning
NALV	Nederlands Actieplan Luchtvaartveiligheid
NLR	Koninklijk Nederlands Lucht- en Ruimtevaartcentrum
NLVP	Nederlands Luchtvaartveiligheidsprogramma
NPAS	Netherlands Plan for Aviation Safety
NVA	Nationale veiligheidsanalyse
OEM	Original Equipment Manufacturer
OVV	Onderzoeksraad voor Veiligheid
SMS	Safety Management System
SSP	State Safety Programme

1 Introductie

1.1 Achtergrond

Beheersing van veiligheidsrisico's is één van de pijlers in het Nederlandse Luchtvaartveiligheidsprogramma (NLVP) [1]. De nationale veiligheidsanalyse (NVA) is daarin een belangrijke schakel. Over de NVA wordt het volgende geschreven in de NLVP: *“Het nationale veiligheidsdoel is het continu verbeteren van de luchtvaartveiligheid door het kennen van de grootste nationale risico's en deze te beheersen tot een acceptabel niveau. Om dit doel te realiseren voert IenW periodiek een veiligheidsanalyse uit. Deze bestaat uit het identificeren van potentiële gevaren, risico's en mogelijke trends voor de hele luchtvaartsector [...] en het formuleren van maatregelen [...]. De nationale veiligheidsanalyse richt zich op de risico's die niet op het niveau van een individuele luchtvaartorganisatie te beheersen zijn, maar een samenwerking tussen luchtvaartorganisaties en overheden vereisen. Minimaal eens per vijf jaar publiceert het ministerie het resultaat in het NALV. [...] IenW neemt in het NALV voor de drie domeinen de grootste risico's op, bijvoorbeeld een top 5, die beheersing op nationaal niveau vereisen.”*

Om de NVA te kunnen uitvoeren is een objectieve informatiepositie nodig op basis van data. De NVA staat niet op zichzelf, maar hangt samen met andere activiteiten om de luchtvaartveiligheid te verbeteren. De NVA heeft een belangrijke relatie met de activiteit 'veiligheidsborging'. Deze activiteit omvat het vaststellen van (kwantificeerbare) indicatoren en streefwaarden voor luchtvaartveiligheid onder meer op nationaal niveau en het monitoren van deze indicatoren. Het NLVP beschrijft de relatie tussen de NVA en veiligheidsborging als volgt: *“Het Ministerie van Infrastructuur en Waterstaat (IenW) vertaalt de uitkomsten van de nationale veiligheidsanalyse naar veiligheidsprestatieindicatoren. Met deze slimme set van indicatoren monitort IenW zowel de ontwikkeling van de risico's als ook de effectiviteit van de maatregelen.”* [1].

1.2 Doel van het onderzoek

In het kader van het Kennis Voor Beleid (KVB)-programma State Safety Plan, onderwerp “Monitoren, analyseren en rapporteren van integrale veiligheid”, heeft het NLR in 2020 een onderzoek uitgevoerd naar het gebruik van veiligheids- en operationele data ten behoeve van de nationale veiligheidsanalyse. Het doel van het onderzoek was het identificeren en het beoordelen van databronnen die gebruikt kunnen worden voor de nationale veiligheidsanalyse, zowel voor gevaarsidentificatie, risicobeoordeling en risicomitigatie. Daarnaast was het doel in kaart te brengen welke verrijkingen en normalisering mogelijk zijn. Tenslotte had het onderzoek als doel om de mogelijkheden te verkennen van Machine Learning technieken om de NVA te ondersteunen.

1.3 Scope van dit onderzoek

De inventarisatie van databronnen is beperkt tot de scope van de (NVA) en deze is als volgt gedefinieerd voor dit onderzoek:

- De NVA richt zich op de risico's die niet op het niveau van een individuele luchtvaartorganisatie te beheersen zijn, maar een samenwerking tussen luchtvaartorganisaties en overheden vereisen. Het zal dan naar

verwachting gaan over een risico dat service provider-overstijgend of systeem-breed is, of interacties tussen partijen betreft.

- Luchtvaartveiligheid is de staat van een luchtvaartstelsel waarin de risico's gerelateerd aan de vliegoperatie zijn gereduceerd en gecontroleerd tot een acceptabel niveau. Luchtvaartveiligheid omvat risico's die betrekking hebben op de operatie van een vliegtuig tussen de tijd dat een persoon aan boord gaat met de intentie te gaan vliegen tot de tijd dat iedereen van boord is [2].
- Gevaren, en het bijbehorende risico, betreffen de luchtvaartveiligheid, waar de overheid op enige manier invloed op kan uitoefenen door middel van wetgeving, beleid, toezichtsactiviteiten, veiligheidspromotie etc. Gevaren en risico's gaan vaak over de operatie, denk daarbij aan voorvallen zoals gecategoriseerd met de CICTT occurrence categorieën. De oorzaken van deze gevaren kunnen dieper in het systeem liggen, zoals het onvoldoende functioneren van een element in het systeemmodel, bijvoorbeeld het toezicht of de 'reporting culture' binnen een organisatie. We nemen aan dat onderliggende gevaren (latente factoren) ook relevant zijn voor de NVA. Data en informatie over de vier elementen van een safety management systeem, waaronder safety risk assessment en safety assurance, vallen binnen de scope.
- Het Nederlandse luchtvaartstelsel omvat het gehele Nederlandse koninkrijk inclusief Bonaire, Sint-Eustatius en Saba (BES-eilanden). Paragraaf 1.4 licht het model toe dat we gebruiken om het Nederlandse luchtvaartstelsel te beschrijven.
- De NVA beperkt zicht tot de drie domeinen binnen het Nederlandse luchtvaartstelsel, te weten: Commerciële luchtvaart, kleine luchtvaart, en onbemande luchtvaart.
- Militaire luchtvaart valt in principe buiten de scope. Er is een raakvlak tussen de NVA en de militaire luchtvaart waar het burgermedegebruik en civiel-militaire samenwerking zoals in de luchtverkeersleiding betreft. Deze twee onderwerpen vallen daarom wel in de scope van de NVA.
- Arbeidsveiligheid, externe veiligheid (of omgevingsveiligheid) en security vallen buiten de scope.

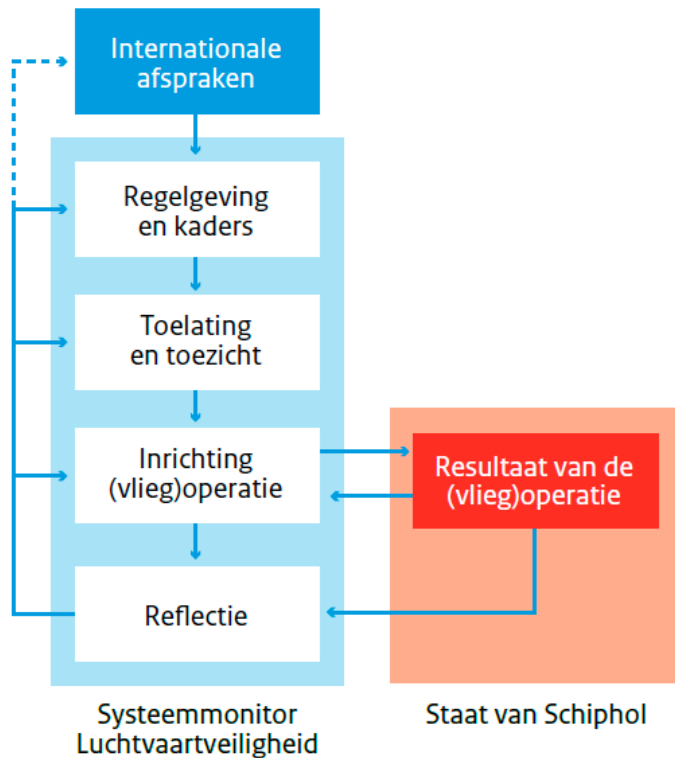
1.4 Het Nederlandse luchtvaartstelsel

Het Nederlandse luchtvaartstelsel betreft alle elementen die een rol spelen voor de Nederlandse luchtvaart.

Vanzelfsprekend omvat het systeem de sectorpartijen die op enige manier hun diensten verlenen binnen Nederland (inclusief de bijzondere gemeentes Bonaire, Sint-Eustatius en Saba (BES-eilanden)). Dit betreft o.a. luchthavens, luchtverkeersleiding en luchtvaartmaatschappijen. Verder omvat het luchtvaartstelsel ook partijen die verder weg staan van de operatie, waaronder de overheid, toezichthouder, en onderzoeksinstanties. Het luchtvaartstelsel is beschreven in het systeemmodel, die ontwikkeld is in het kader van de systeemmonitor [3]. Figuur 1 toont de zes elementen die elk op hun eigen manier invloed hebben op de veiligheid van de Nederlandse luchtvaart.

- Internationale afspraken* betreft alle regels en standaarden die zijn gesteld door internationale organisaties zoals ICAO en EASA.
- Op nationale niveau stelt de Nederlandse overheid *regelgeving en kaders* op. Verder bepaalt de Nederlandse overheid het beleid en de begroting voor o.a. handhaving en toezicht.
- Toelating en toezicht* betreft de daadwerkelijke uitvoering hiervan door ILT. Denk hierbij aan certificeringsaanvragen en audits.
- De *inrichting van de vliegoperatie* betreft alle organisatorische inrichtingen die de vliegoperatie ondersteunen. Denk hierbij aan veiligheids- en kwaliteitsmanagement van de sectorpartijen, het ISMS, enzovoorts. De inrichting maakt het mogelijk om daadwerkelijke de (vlieg-)operatie tot uitvoering te brengen.
- Het *resultaat van de (vlieg-)operatie* betreft de effecten die de operatie heeft op veiligheid, financiën, netwerk, milieu, enzovoorts.

- f. *Reflectie* sluit de cirkel en zorgt ervoor dat de resultaten van de vliegoperatie gebruikt worden om te leren en bij te sturen binnen het systeem. Voorbeelden van reflectie zijn onderzoeken van de OVV en (integrale) veiligheidsanalyses.



Figuur 1: Overzicht van de systeemmonitor, overgenomen uit het NLVP [1].

1.5 Leeswijzer

In hoofdstuk 2 worden de resultaten van de inventarisatie van beschikbare databronnen beschreven. Hoofdstuk 3 gaat in op de mogelijke verrijking en normalisatie van veiligheidsdata met contextdata. In hoofdstuk 4 beschrijven we het literatuuronderzoek dat is gedaan naar verschillende Machine Learning technieken die mogelijk toegepast kunnen worden in een NVA. Het hoofdstuk gaat ook in op de toepassing van deze technieken in een drietal cases. Hoofdstuk 5 bevat de conclusies en aanbevelingen.

2 Databronnen

2.1 Inleiding

Het uiteindelijke doel van de NVA is het beheersen van de grootste risico's voor het Nederlandse luchtvaartstelsel tot een acceptabel niveau. Hiervoor is het niet afdoende om alleen de gevaren in kaart te brengen. Om effectief de risico's te beheersen is het van belang om te weten hoe groot de risico's van deze gevaren zijn om daarmee prioriteiten aan te kunnen brengen. Daarnaast is het essentieel om te begrijpen waarom de risico's bestaan en in welke context deze risico's zich manifesteren. Dat begrip is belangrijk om effectieve en efficiënte maatregelen te kunnen implementeren om de risico's te beheersen. De identificatie van databronnen is gestart vanuit het oogpunt van de volgende vier vragen die de NVA moet beantwoorden, waarbij databronnen deze analyse moeten ondersteunen.

Wat zijn de gevaren in de Nederlandse luchtvaart voor de drie domeinen: CAT, GA en drones?

Input voor het beantwoorden van deze vraag kan komen uit diverse bronnen, zoals voorvalmeldingen, statistieken (bijvoorbeeld uit jaarverslagen, dashboards), of uit analyses waarin reeds duiding is gegeven aan risico's (bijvoorbeeld in veiligheidsstudies, actieplannen, workshops met experts).

Wat zijn de grootste risico's voor de drie domeinen: CAT, GA en drones?

Gevaren dienen geprioriteerd te worden op basis van risico, d.w.z. kans en impact van het gevaar. Dit vraagt o.a. (statistische) analyses, waarbij contextuele data voor zowel normalisatie van voorvallen als achtergrondinformatie belangrijk zijn om risico's te kunnen duiden.

Hoe komt het dat deze risico's in het systeem bestaan? Welke factoren spelen een rol?

Om de risico's effectief en efficiënt te kunnen mitigeren, moet men enig begrip hebben van de oorzaken en factoren die een rol spelen in het ontstaan van het risico. Denk hierbij aan de operationele context (bijvoorbeeld weerscondities, baangebruik), context van de organisatie (bijvoorbeeld effectiviteit van het SMS, de safety culture), of (inter)nationale context (b.v. wetgeving, effectiviteit van toezicht).

Welke maatregelen zijn nodig om de risico's, gegeven de context waarin ze zitten, te beheersen?

Beheersmaatregelen kunnen aangrijpen op de operatie-, organisatie- en (inter)nationale wet- en regelgeving (de kaders). Databronnen kunnen bijdragen aan de identificatie en onderbouwing van beheersmaatregelen, en bij het beoordelen van de verwachte effectiviteit van maatregelen. Na invoering van maatregelen kan men met behulp van data monitoren in hoeverre de genomen maatregelen effectief zijn.

2.2 Aanpak

Voor het opstellen van de lijst met databronnen is de voorbeeldlijst in het NLVP [1] als inspiratie gebruikt, aangevuld met ervaringen en kennis van NLR-collega's. Tevens is een beoordelingskader geformuleerd waarmee de databronnen worden gecategoriseerd op verschillende onderwerpen. Dit maakt belangrijke karakteristieken van de databronnen inzichtelijk en geeft een eerste indruk van de compleetheid, toegankelijkheid, dekkingsgraad en reikwijdte van de databronnen.

Het beoordelingskader is als volgt gedefinieerd:

- Informatietype:
 - Hazarddata: data over mogelijke voorvallen, risico's, gevaren (hazards) en bedreigingen (threats).
 - Contextdata: data over de achtergrond en omstandigheden van de operatie en het luchtvaartstelsel.
 - Beleidsdata: data over wet- en regelgeving, (inter)nationaal beleid, kaders voor het Nederlandse luchtvaartstelsel.
- Analyseniveau:
 - Ruwe data: beschrijft niet geïnterpreteerde losse datapunten (zoals radar- en weersdata), tekstdata (zoals meldingen), of handboeken.
 - Statistiek: samenvattingen van data zonder interpretatie, statistiek van data.
 - Analyse: (trend)analyse en duiding van data, verklaring en toelichting op oorzaken en gevallen, omvat interpretaties en conclusies.
 - Mitigatie: rapportages waarin verbetermaatregelen of aanbevelingen beschreven zijn.
- Datatype: Document of dataset (getallen, cijferreeksen e.d.).
- Eigenaar, geeft aan welke partij de eigenaar en/of beheerder van de databron is.
- Type eigenaar, geeft aan:
 - Nederlandse overheid.
 - Nederlandse luchtvaart sectorpartij.
 - Europese overheid.
 - Internationale instantie.
 - Kennisinstituut/universiteit/onderzoeksinstituut.
 - Overig.
- Toegankelijkheid, geeft aan of data vrij toegankelijk (open source) of in eigendom van een partij is:
 - Open-source: Data die vrij en gratis toegankelijk is voor iedereen.
 - Open-source (Rijksoverheid): Dit is een subgroep van open-source data, die eigendom is van het Nederlandse Rijk. De aanname is dat deze data makkelijk toegankelijk is voor de NVA.
 - Proprietary: Data die in eigendom is van een derde partij (niet overheid). Deze data is mogelijk onder voorwaarden te verkrijgen voor de NVA.
 - Proprietary (Rijksoverheid): Dit is een subgroep van proprietary data, die eigendom is van het Nederlandse Rijk. Aanname is dat deze data makkelijk toegankelijk is voor de NVA.
- Domein, geeft aan over welk domein de databron informatie bevat:
 - Commerciële luchtvaart (Commercial Air Transport (CAT)).
 - Kleine luchtvaart (General Aviation (GA)).
 - Onbemande luchtvaart (drones).
- Service provider, beschrijft over welke partij of dienstverlener in de luchtvaart de databron informatie bevat:
 - Airport
 - ANSP: Air Navigation Service Provider.
 - Drone: Drone organisatie.
 - GA: General Aviation (GA) organisatie.
 - Ground: Ground handling services (grondafhandelaar).
 - Mx: Maintenance organisation (onderhoudsbedrijf).
 - OEM: Original Equipment Manufacturer.
 - Operator.
 - TO: Training organisation.
- Systeemmodel (paragraaf 1.4), beschrijft over welk element van het systeemmodel de databron informatie bevat:
 - A: Internationaal afspraken.
 - B: (Nationale) Regelgeving en kaders.

- C: Toelating en toezicht.
- D: Inrichting (vlieg)operatie.
- E: Resultaat van de (vlieg)operatie.
- F: Reflectie.

2.3 Resultaten

2.3.1 Lijst met databronnen

Het resultaat van de inventarisatie van databronnen is de lijst met 57 databronnen die geclassificeerd zijn volgens het beoordelingskader uit de vorige paragraaf. De lijst is opgenomen als bijlage bij dit rapport, zie Appendix A. De volgende paragrafen beschrijven enkele karakteristieken van de databronnen in meer detail.

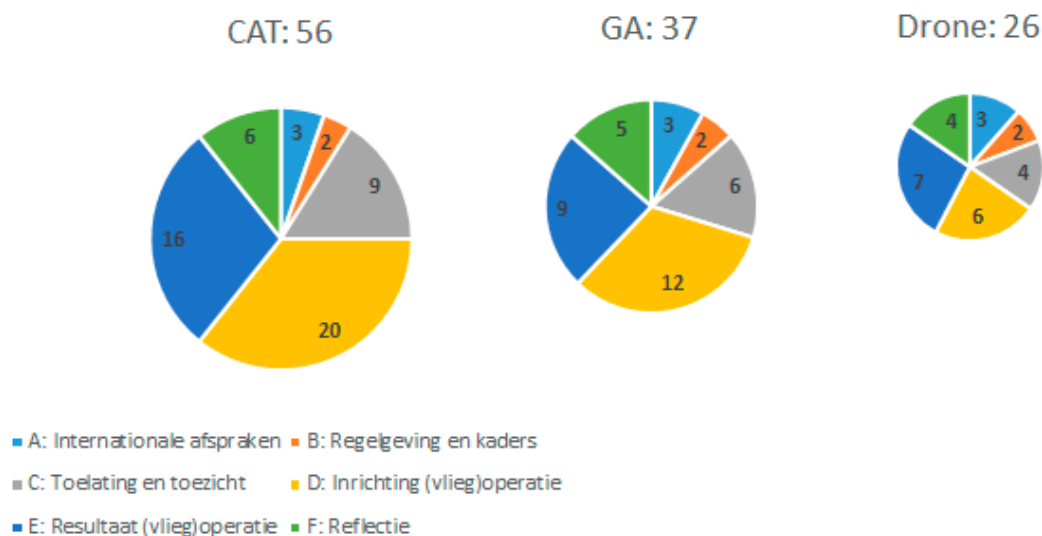
2.3.2 Kwaliteit

Goede kwaliteit data is essentieel voor de NVA en de beoordeling van data(bronnen) zal onderdeel zijn van de uitvoering van de NVA. De inzet van analysemethodieken hangt namelijk mede samen met de kwaliteit van de beschikbare data. Paragraaf 2.4 beschrijft de weging van databronnen en mogelijke criteria om de kwaliteit van databronnen te beoordelen. Daarnaast worden enkele adviezen gegeven voor het gebruik van verschillende soorten databronnen in de NVA. Er is in dit onderzoek geen beoordeling uitgevoerd van de kwaliteit van de geïdentificeerde databronnen.

2.3.3 Dekkingsgraad

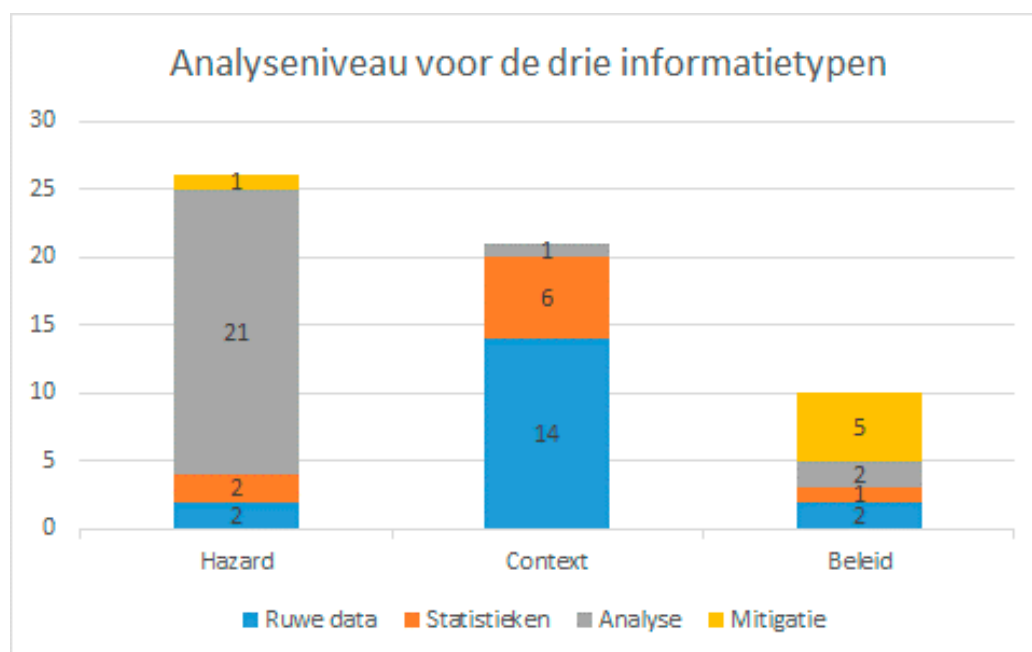
Figuur 2 laat zien hoe de databronnen verdeeld zijn over de drie domeinen (CAT, GA en drone) en over de verschillende elementen van het systeemmodel. Een databron kan meerdere domeinen afdekken en is toegekend aan één element van het systeemmodel. De meeste databronnen (56 bronnen) betreffen de commerciële luchtvaart, gevolgd door kleine luchtvaart (37 bronnen) en onbemand (26 bronnen). Circa driekwart van de databronnen bevat informatie over de systeemelementen inrichting (D), het resultaat van de (vlieg)operatie (E), en de reflectie (F). De systeemelementen internationale afspraken en regelgeving (A en B) en toezicht (C) worden in mindere mate afgedekt door de geïdentificeerde databronnen.

De dekkingsgraad van databronnen wordt mede beïnvloed door de volwassenheid van beleid en regelgeving, en het veiligheidsmanagementsysteem van de partijen. Voor enkele onderwerpen (bijvoorbeeld, drones en de BES-eilanden) is gestructureerde dataverzameling nog in ontwikkeling. Bij het beoordelen van de dekkingsgraad is gekeken in hoeverre informatie over een onderwerp in potentie aanwezig zou kunnen zijn in de databronnen, bijvoorbeeld binnen het domein drone. Incidenten met drones kunnen bijvoorbeeld aan het ABL gemeld worden. Echter, er is in deze studie niet onderzocht in hoeverre daadwerkelijk op dit moment de databron ook kwalitatief goede informatie bevat over een bepaald onderwerp. Er is ook geen onderzoek gedaan naar de aanwezigheid van informatie in databronnen over het luchtvaartstelsel op de BES-eilanden.



Figuur 2: Aantal databronnen voor verschillende onderdelen in het systeemmodel, per domein weergegeven.

Figuur 3 toont de verschillende type informatie in de geïdentificeerde databronnen. De 57 databronnen zijn onder te verdelen in 26 databronnen van het informatietype “hazarddata”, 21 bronnen “contextdata”, en 10 bronnen “beleidsdata”. Van de 26 bronnen met hazarddata bevatten 4 bronnen ruwe data (bijvoorbeeld ABL-meldingen en meldingen van sectorpartijen) en statistieken (bijvoorbeeld ABL dashboard). De overige 22 databronnen bevatten informatie over analyses en mitigaties, zoals OVV-rapporten en veiligheidsanalyses. De 21 databronnen met contextdata bevatten veel ruwe data en statistieken, en geven informatie over de bijvoorbeeld de omstandigheden waarin de operatie wordt uitgevoerd. De 10 databronnen met beleidsdata bevatten met name wet- en regelgeving, (inter)nationaal beleid, en kaders voor het Nederlandse luchtvaartstelsel.



Figuur 3: Informatietype en analyseniveau van de databronnen.

Databronnen die informatie leveren over statistieken, analyses en mitigaties bevatten meestal informatie die gebaseerd is op (onderliggende) ruwe data. Tabel 1 geeft aan welke ruwe databronnen gebruikt worden in de

databronnen die aangemerkt zijn met analyseniveau “statistiek”, “analyse” en “mitigatie”. Deze databronnen zijn gegroepeerd in een aantal thema’s: Beleidseffectiviteit, Beleidsstukken, Toezichtsrapportages (ILT), ISMS Roadmap, Overlegorgaan, SMS effectiviteit, Veiligheidsanalyses en veiligheidsprestatie, voorvalonderzoeken. Per thema zijn een aantal kenmerkende publicatie of databronnen genoemd als voorbeeld, met in de rechterkolom een opsomming van typische onderliggende databronnen.

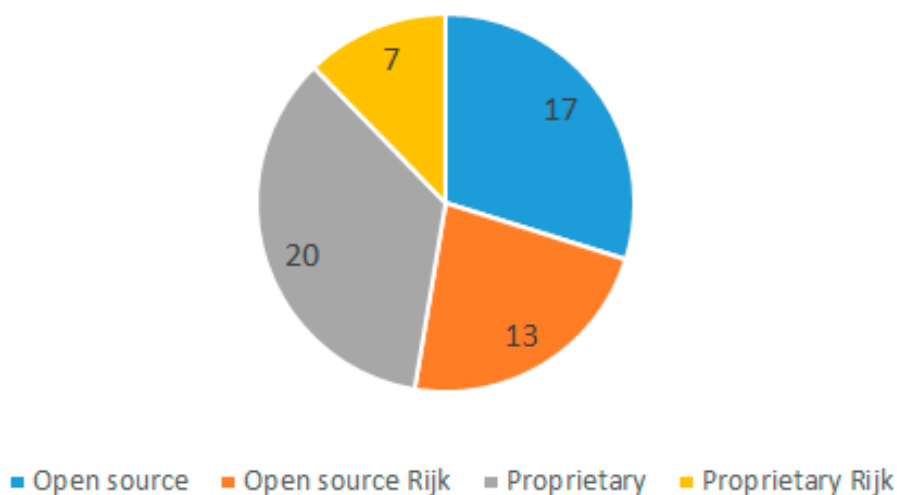
Tabel 1: Onderliggende ruwe data in databronnen met statistiek, analyse en mitigatie.

Databron met statistiek, analyse en mitigatie	Welke publicaties vallen hieronder? Voorbeelden:	Onderliggende data (niet volledig)
Beleidseffectiviteit	EU Performance Scheme (Reference Periods), ILT Jaarverslag, Landelijke Systeem Monitor, Toezichtsprogramma (Schiphol/Luchtvaart), Jaarverslagen luchthavens	<ul style="list-style-type: none"> • Voorvalmeldingen ABL (EU 376/2014) • Voorvalmelding sector intern • Vliegbewegingen/ADS-B en radar data • Procedures (safety management, operation manuals) • Baangebruik Schiphol
Beleidsstukken	Europees EASP, Europees EPAS, Internationaal GASP, NALV, NLVP	<ul style="list-style-type: none"> • Voorvalmeldingen EU (EU 376/2014)
Toezichtsrapportages (ILT)	Handhavingsrapportages ILT, resultaten audits en inspecties, Signaalrapportages ILT	<ul style="list-style-type: none"> • Voorvalmeldingen ABL (EU 376/2014) • Inspectie en auditresultaten • Procedures (safety management, operation manuals)
ISMS Roadmap	ISMS Roadmap, evaluatie effectiviteit ISMS maatregelen	<ul style="list-style-type: none"> • Voorvalmelding sector intern • Vliegbewegingen/ADS-B en radar data • Weersdata • Baangebruik Schiphol (baancombinaties, baankruisingen)
Overlegorgaan	Aanbevelingen expertgroep luchtvaart, EASA Collaborative Analysis Groups (CAGs), notulen ABL CA+ overleg, ISMS TopPSAG, Network of Analysts overleggen	<ul style="list-style-type: none"> • Voorvalmeldingen ABL (EU 376/2014) • Voorvalmelding sector intern • Vliegbewegingen/ADS-B en radar data
SMS effectiviteit	Externe evaluatie ABL, externe evaluatie ISMS, ILT audits	<ul style="list-style-type: none"> • Inspectie en auditresultaten • Voorvalmelding sector intern • Procedures (safety management, operation manuals)
Veiligheidsanalyses en veiligheidsprestatie	Veiligheidsanalyses door derde partijen, Integrale Veiligheidsanalyses (IVA), ILT publicatie Staat van Schiphol, publicatie top-5 ground/flight safety hazards/risks sector ISMS Performance Review Report, ANS Performance,	<ul style="list-style-type: none"> • Voorvalmeldingen ABL (EU 376/2014) • Procedures (safety management, operation manuals) • Voorvalmelding sector intern • Vliegbewegingen/ADS-B en radar data
Voorvalonderzoeken	ABL factsheets, ABL dashboard luchtvaartvoorvallen, EASA Annual Safety Review, EVAIR Safety Report, LVNL Overzicht Voorvallen, Onderzoeken en kwartaalrapportage OVV	<ul style="list-style-type: none"> • Voorvalmeldingen ABL (EU 376/2014) • Procedures (safety management, operation manuals) • Voorvalmelding sector intern • Vliegbewegingen/ADS-B en radar data • Vliegbewegingen/ADS-B en radar data

2.3.4 Eigendom en toegankelijkheid

Figuur 4 laat zien wat de toegankelijkheid is van geïdentificeerde databronnen op basis van eigenaarschap van de data. Ongeveer de helft van de databronnen is vrij toegankelijk (open source) of eigendom van de Nederlandse overheid (proprietary rijk) en kan daarom relatief eenvoudig beschikbaar gemaakt worden voor de NVA. In Tabel 2 is aangegeven welke bronnen van het Rijk als proprietary zijn aangemerkt. Twintig bronnen zijn aangemerkt als proprietary, dus niet openbaar en in eigendom van andere partijen dan het Rijk. Deze databronnen zijn bijvoorbeeld operationele data (b.v. radardata en data over baangebruik), veiligheidsmeldingen in rapporteringsystemen van sectorpartijen, (interne) risicoanalyses en bijvoorbeeld gespreksverslagen van overlegorganen (bijvoorbeeld ISMS TOPSAG en Network of Analysts (NOA)).

Toegankelijkheid databronnen



Figuur 4: Toegankelijkheid van de databronnen.

Tabel 2: Proprietary databronnen, eigendom van het Rijk.

Analyseniveau	Databron
Ruwe data	Voorval meldingen ABL
Statistieken	ABL dashboard luchtvaartvoorvallen
Analyse	<ul style="list-style-type: none"> Resultaten ILT audits en inspecties Meeting notes ABL Meetings CA+ Meeting notes ABL Meetings GA+
Mitigatie	<ul style="list-style-type: none"> Integrale veiligheidsanalyses Aanbevelingen expertgroep luchtvaart

2.4 Kwaliteit van data en waarde voor de NVA

2.4.1 Achtergrond

De beschikbaarheid van gegevens van goede kwaliteit is essentieel voor het uitvoeren van een nationale veiligheidsanalyse. Het gaat hierbij niet alleen om specifieke veiligheidsgegevens zoals voorvalmeldingen, incidentgegevens, resultaten van ongevallenonderzoek en dergelijke, maar ook om contextuele gegevens zoals aantallen vluchten, weersinformatie of vlootgegevens. De kwaliteit van de databronnen moet kunnen worden beoordeeld en vastgelegd. Er moeten daarnaast richtlijnen zijn voor de combinatie van gegevens uit verschillende bronnen, zeker als de kwaliteit van de bronnen verschillend is.

2.4.2 Kwaliteit van databronnen

De kwaliteit van de data(bronnen) wordt bepaald door de mate waarin data actueel, nauwkeurig, betrouwbaar, representatief, en volledig is en voldoende informatie en detailniveau bevat om een kwalitatief goede analyse uit te kunnen voeren. Deze kwaliteitscriteria zijn in de onderstaande Tabel 3 beschreven. De analist zal een databron op deze criteria kunnen beoordelen. Er zijn een aantal vragen geformuleerd, als leidraad voor de analist, om de kwaliteit van de data(bron) voor het betreffende criterium te beoordelen.

Tabel 3: Overzicht mogelijke kwaliteitscriteria voor databronnen.

Kwaliteitscriteria	Definitie en leidraad
Actualiteit	Mate van hoe actueel de beschikbare data is. Welke tijdsperiode beslaat de data? Hoe recent of oud is de data?
Nauwkeurigheid	Mate waarin de data nauwkeurig of exact is. Hoe precies of accuraat is de data? Wat is de fout(marge) bij de meting en in de data? Hoe goed wordt de werkelijkheid 'gemeten' of 'weergegeven'?
Betrouwbaarheid	Mate van hoe statistisch betrouwbaar of geloofwaardig de data is. Wat is de statistische betrouwbaarheid van data of steekproef? Zijn de betrouwbaarheidsintervallen bekend? In welke mate is de data te controleren op herkomst en feiten? Welk deel van de data(bron) bevat objectieve of subjectieve data? Wat is het aandeel objectieve/subjectieve data? Wat zijn de kwalificaties van de rapporteur (in geval van subjectieve data)?
Representativiteit	Mate waarin data representatief is voor het vraagstuk dat geanalyseerd wordt. In hoeverre dekt de data(bron) de volgende zaken af: het Nederlandse luchtvaartstelsel, de onderdelen van het systeemmodel, de CAT/GA/drone domeinen, de huidige of toekomstige situatie, de uitvoering van de operatie, de infrastructuur (grond en lucht), niveau van kennis en competenties, training, organisatie aspecten etc.
Volledigheid	Mate waarin data compleet is. In hoeverre zijn datavelden 'gevuld' met informatie? Hoeveel informatie ontbreekt in de databron? Wat is de mate van underreporting of overreporting (doublures)? Wat is de dekkingsgraad van de data?
Informatiedichtheid	Mate waarin databron informatie biedt en het detailniveau van informatie. Wat is het detailniveau, diepgang, breedte (scope) van de data? In hoeverre biedt de data inzicht in oorzaken, gevolgen, factoren? Tot op welk detailniveau wordt informatie beschreven in de databron? In welke mate is de data gecategoriseerd of gelabeld?

Kwaliteitscriteria	Definitie en leidraad
Valide	Mate waarin de data van een "test" en het te meten verschijnsel overeenkomen ("meet de test wat hij zou moeten meten").
Continuïteit	Mate waarin de data actueel aangevuld wordt. Hoe vaak wordt deze data bijgewerkt? Wanneer is deze data voor het laatste bijgewerkt? Wordt de data real time of near real time verzameld en verwerkt? Is er sprake van een continue of onderbroken stroom in datavergaring? In hoeverre is toekomstige data van deze databron gegarandeerd?

2.4.3 Waarde van databronnen voor de NVA

In de beschrijving van de waarde van databronnen voor de NVA maken we onderscheid tussen objectieve en subjectieve data:

Objectieve data	Subjectieve data
Systeemdata	Meldingen
Analyserapporten	Observaties
Ongevalsrapporten	Expertmening

Objectieve data – Systeemdata

Databronnen die data of informatie bevatten die verkregen zijn uit technische systemen en sensoren leveren over het algemeen veel data van hoge kwaliteit. Voorbeelden zijn vluchtdata geregistreerd door bijvoorbeeld het Aircraft Condition Monitoring System (ACMS), Flight Data Monitoring (FDM) events data, radar track data, ADS-B data, en meteogegevens (bijvoorbeeld METAR). Contextdata zoals aircraft performance data, vlootdata, of procedures zijn in hoge mate objectief en goed meetbaar vast te stellen.

De systeemdata geven de systeemtoestand (*system state*) goed weer en bieden contextinformatie voor analyses. Het nadeel is dat dergelijke systeemdata vaak niet 'het verhaal er achter vertellen'. Informatie over menselijke waarneming, interpretatie, besluitvorming en acties (de *human factors*) van het operationeel personeel, en organisatorische factoren worden niet goed vertegenwoordigd in deze databronnen. Data gemeten door systemen geven over het algemeen – in vergelijking met meldingen/observaties – veel meer informatie, zijn onafhankelijker en objectiever. De nauwkeurigheid, betrouwbaarheid, volledigheid, informatiedichtheid, validiteit en continuïteit, van systeemdata zijn vaak beter dan bij zelfrapportages. Systeemdata kunnen fouten bevatten (bijv. corrupte data, ontbrekende data, meetfouten), maar men kan in de dataverwerking en analyse deze fouten herstellen, inzichtelijk maken, en de betrouwbaarheid van het resultaat vaststellen.

Objectieve data – Analyserapporten

Analyserapporten en ongevalsrapporten bevatten vaak een combinatie van objectieve en subjectieve data. Hoewel over het algemeen objectieve data ten grondslag liggen aan deze rapporten, wordt de interpretatie van deze data en daarmee de analyse in zekere mate beïnvloed door de kennis, ervaring, visie, en inzichten van analisten en onderzoekers. Hoewel subjectieve data, zoals meldingen en expert mening, verwerkt kunnen zijn in analyserapporten, scharen we deze analyserapporten onder objectieve data gelet op beoordeling, weging en onderbouwing die wordt uitgevoerd als onderdeel van de analyse. Men mag namelijk verwachten dat analyserapporten informatie verschaffen over de onderbouwing van de analyse, conclusies en aanbevelingen. De onderliggende databronnen, de beoordeling van de kwaliteit van deze data, en de weging van deze bronnen zullen in een kwalitatief goed rapport worden beschreven. De analist kan het effect van de kwaliteit van de data op de uitkomst van de analyse beoordelen op

basis van de informatie uit het rapport, en daardoor een inschatting maken over de bruikbaarheid en waarde van het analyserapport voor de NVA. Men kan in het algemeen stellen dat analyserapporten en themastudies over veiligheidsonderwerpen opgesteld door onderzoeks- en kennisinstituten (bijv. NLR, DLR, NASA), industrie en brancheorganisaties (bijv. Airbus, Boeing, ICAO, IATA, FSF, CANSO, Eurocontrol), en toezichthouders (bijv. FAA, EASA) van goede kwaliteit zijn.

Objectieve data – Ongevalsrapporten

Ongevalsrapporten zijn door een onafhankelijke onderzoeksraad (bijv. de OVV) opgesteld. Voor dergelijke rapporten gelden internationale standaarden en aanbevelingen (ICAO Annex 13). Over het algemeen zijn de conclusies en aanbevelingen in dergelijke rapporten goed onderbouwd met data uit verschillende bronnen, die gewogen en geanalyseerd zijn. Daarnaast stelt een onderzoeksraad betrokken partijen in de gelegenheid reviewcommentaar te leveren op het (concept)rapport. Deze zaken dragen bij aan de kwaliteit. De kanttekening is wel dat een ongevalsrapport zich beperkt tot één specifiek voorval en dat het benutten van de informatie uit dergelijke rapporten in een bredere context met enige voorzichtigheid moet worden gedaan.

Subjectieve data – Meldingen

Verplichte en vrijwillige meldingen van onveilige situaties en incidenten worden beïnvloed door (subjectieve) waarneming en meldingsbereidheid. De meldingsbereidheid kan beïnvloed worden door bijvoorbeeld de actualiteit, safety/just culture, en gebruiksgemak van rapportagesystemen. Het is bekend dat underreporting in meldingen aanwezig is, maar de mate van underreporting is over het algemeen niet bekend. Er zijn diverse studies naar dit onderwerp uitgevoerd. Het beeld dat naar voren komt is dat gemiddeld 50-60% van de verplichte meldingen ook daadwerkelijk gerapporteerd wordt. De meldingsgraad kan variëren van 1% tot 85% afhankelijk van het onderwerp [4]. Meldingen over technische systemen, onderhoud, luchtwaardigheid, en waarvoor een verplichting bestaat (EU-verordening 2014/376) zijn over het algemeen betrouwbaarder, dan (vrijwillige) meldingen over operationele voorvallen. Het is bekend dat de mate van underreporting hoger is daar waar het gaat om het melden van gevoelige zaken en eigen fouten.

De meldingsbereidheid kan hoog zijn op een bepaald onderwerp wanneer het onderwerp de aandacht heeft in het bedrijf, onder personeel of in de media. Daarentegen kan de meldingsbereidheid afnemen als het personeel het gevoel heeft dat er niks met de meldingen wordt gedaan en 'meldingsmoe' wordt. De meldingsbereidheid kan ook teruglopen omdat het personeel denkt dat het probleem wel bekend is bij het management en er maatregelen worden genomen, waardoor de noodzaak om te (blijven) rapporten minder gevoeld wordt. Deze aspecten (actualiteit, meldingsmoeheid, gevoelde noodzaak) kan de analist meewegen bij het inschatten van de mate van underreporting over een onderwerp.

De kwaliteit van de meldingen is heel divers wat betreft informatiedichtheid, representativiteit, betrouwbaarheid, volledigheid en actualiteit. De kwaliteit van de meldingen moet daarom echt per casus worden vastgesteld door de analist.

Subjectieve data – Observaties

Informatie uit audits, inspecties, en lijnobservaties (LOSA) is ook gevoelig voor waarneming en subjectiviteit. De kwaliteit van deze informatie is divers en moet daarom per casus worden vastgesteld door de analist. Het voordeel ten opzichte van meldingen is dat deze informatie op een gestandaardiseerde manier wordt verzameld en dat de informatie aangeleverd wordt door specifiek daarvoor getraind personeel (inspecteurs en auditors), wat bijdraagt aan kwaliteit.

Expertmening

Het inzetten van expertmening is mogelijk voor bijvoorbeeld het identificeren van (nieuwe) gevaren en het inschatten van het risiconiveau. Het soort en de mate van expertise die nodig is, is afhankelijk van de casus. Het inzetten van expertmening kan op een gestructureerde manier worden uitgevoerd, wat bijdraagt aan de kwaliteit van deze 'data'. Er zijn methodes beschikbaar om de experts onderling te kalibreren, wat relevant is voor het wegen van de meningen van verschillende experts [5,6]. Daarnaast moet de uitvraag worden gestructureerd volgens een protocol om op een uniforme wijze de uitvraag onder experts uit te voeren.

2.4.4 Combineren van databronnen voor de NVA

De volgende adviezen kunnen worden gegeven met betrekking tot het combineren van verschillende soorten data(bronnen):

- Er zijn verschillende databronnen en soorten data beschikbaar voor de analyse in de NVA. Het soort en de kwaliteit van de data die beschikbaar zijn heeft invloed op het type analyse en de analysetechnieken die worden ingezet. Zowel de kwaliteit van de data speelt een rol als de mate waarin de analist in staat is met/zonder analysetechnieken informatie uit de data te halen
- Afhankelijk van de casus zal de analist moeten beoordelen welke databronnen geschikt zijn en hoeveel waarde aan een bron kan worden toegekend voor het identificeren van gevaren en het beoordelen van het risiconiveau. Het is uiteindelijk aan de analist om, op basis van zijn/haar kennis en ervaring, een inschatting te maken van de inzet van diverse databronnen, een oordeel te vellen over de kwaliteit van de data, en bronnen ten opzichte van elkaar te wegen in relatie tot de casus. Daarbij zullen ook praktische overwegingen een rol spelen zoals beschikbare tijd en middelen om data te verkrijgen, te verwerken en te analyseren. Het is belangrijk dat een analist zo goed mogelijk beperkingen in de data en de implicaties voor de betrouwbaarheid van de risicobeoordelingen documenteert als onderdeel van de analyse.
- Als objectieve data beschikbaar is, die bovendien goed scoort op de kwaliteitscriteria, dan heeft het gebruik van deze data de voorkeur boven subjectieve data. Objectieve data zijn over het algemeen ook uniformer in kwaliteit dan subjectieve data. Wanneer de beschikbare objectieve data van mindere kwaliteit is of wanneer objectieve data ontbreekt, dan kan de analist meer waarde hechten aan subjectieve data.
- Wanneer subjectieve data worden ingezet, zal het beoordelen van de kwaliteit essentieel zijn, omdat deze data(bronnen) vaak worden gekenmerkt door een wisselende kwaliteit. Contextuele data kunnen helpen om de subjectieve data zo goed mogelijk te interpreteren. Denk bijvoorbeeld aan het normaliseren van de subjectieve data. Bijvoorbeeld, het aantal meldingen over een type voorval kan worden afgezet tegen het aantal uitgevoerde vluchten of de weerscondities tijdens het voorval.
- Voor de identificatie van nieuwe of 'toekomstige' gevaren en het beoordelen van risico's waarover geen of weinig objectieve data beschikbaar is, kan men gebruik maken van expertmening. Indien men expertmening inzet voor een risicobeoordeling, wordt aanbevolen om deze experts te kalibreren volgens de methode van [6].
- Informatie over gevaren, de ernst en frequentie van gevaren is gevoeliger voor subjectiviteit dan contextinformatie, omdat het vaak gebaseerd is op waarnemingen (meldingen en observaties) en expertmening.

2.5 Beschouwing

Er is een aanzienlijke hoeveelheid databronnen beschikbaar voor de NVA. Kijkend naar de lijst met databronnen stellen we de volgende zaken vast:

- We hebben een groot aantal databronnen geïdentificeerd met een enorm volume aan data en informatie. De geïdentificeerde databronnen dekken de drie domeinen en de elementen van het systeemmodel af, waardoor deze bronnen kunnen bijdragen aan het vormen van een breed beeld van de stand van de nationale luchtvaartveiligheid in de NVA.
- De dekkinggraad van databronnen voor de domeinen General Aviation (GA) en drone operations is zoals verwacht een stuk lager dan voor Commercial Air Transport (CAT). Er zijn minder databronnen beschikbaar met informatie over GA en drones operaties. Over het algemeen verwachten we hogere kwaliteit en meer informatie in bronnen met betrekking tot CAT ten opzichte van GA en drone operations. De mate waarin de luchtvaartoperaties op de BES-eilanden (Bonaire, Sint Eustatius, Saba) gedekt zijn in de databronnen is onduidelijk. Wel zijn er een aantal bronnen te noemen waar met zekerheid over kan worden gezegd dat deze geen informatie bevatten omtrent de situatie op de BES-eilanden.
- Men kan onderscheid maken tussen de kwaliteit van data en van informatie. Bij informatie gaat het over de betekenis die data heeft voor een bepaald doel, bijvoorbeeld voor de vaststelling van de ernst van een voorval. De kwaliteit van zowel de data als de informatie verschilt per databron. Bijvoorbeeld, radar data is van hoge kwaliteit (nauwkeurig, betrouwbaar etc.) en biedt veel informatie over zowel CAT als GA flight tracks en bijvoorbeeld separatieonderschrijdingen. Meldingen aan het ABL kunnen zowel in het CAT als GA domein van mindere kwaliteit zijn en/of weinig informatie bevatten voor bijvoorbeeld een analyse van oorzaken.
- De sectorpartijen beschikken over databronnen die ruwe, niet geanalyseerde data bevatten, zoals vluchtdata, radardata, en voorvalmeldingen. Deze databronnen zijn groot in volume en variëteit, en het vraagt veel inspanning en middelen (kennis en systemen) om deze data te verwerken, op te slaan, en te analyseren voor een risicoanalyse. Daarom moet in de NVA aanpak uitgewerkt worden voor welke type vraagstukken het nodig en de moeite waard is om deze ruwe data te benutten.
- De overheid (Ministerie, ILT, ABL) beschikt met name over databronnen die statistiek en analyses bevatten. Deze informatie is eenvoudig toegankelijk en laagdrempelig te gebruiken in een NVA gelet op de toegankelijkheid en de geringe inspanning om deze data te ontsluiten, in vergelijking met ruwe data.
- Het gebruik van statistiek en analyses in de NVA is niet altijd 'rechttoe, rechtaan', omdat voor statistiek keuzes zijn gemaakt en interpretaties zijn toegepast. Gegevens uit verschillende statistieken zijn meestal lastig te combineren vanwege de verschillende contexten en dimensies, en onduidelijkheid over overlap van onderliggende datapunten.
- Sectorpartijen voeren ook veel interne risicoanalyses uit, die bruikbaar kunnen zijn voor de NVA. Echter, dergelijke risicoanalyses zijn niet openbaar. Het Ministerie zal dus met sectorpartijen in overleg moeten treden over of en onder welke voorwaarden dergelijke risicoanalyses beschikbaar gemaakt kunnen worden voor de NVA.
- Er zijn veel soorten data beschikbaar in databronnen die benut kunnen worden in de analyse van gevaren en risico's, en context kunnen geven in dergelijke analyses. Op hoofdlijnen biedt deze contextuele data informatie over de mate van activiteit of productie van een partij, en/of over het 'systeem' en de omgeving waarin de operatie plaatsvindt.

3 Dataverrijking en normalisering

3.1 Inleiding

Tijdens de analyse van data kan het zinvol zijn andere, aanvullende informatie of gegevens te gebruiken om de data te verrijken of te normaliseren, om zo effecten van de context weer te geven. Verrijking betekent dat een databron door het toevoegen van andere data(bronnen) 'rijker' aan informatie wordt. Een voorbeeld is het toevoegen van de weerscondities en landingsbaan aan gemelde go-arounds. Normaliseren houdt in dat het aantal observaties, bijvoorbeeld het aantal incidenten, afgezet wordt tegen contextuele factoren, bijvoorbeeld het weer of een type operatie. Bijvoorbeeld, het aantal incidenten per 10.000 vluchten, per maand, of in een bepaalde zichtconditie. Verrijking en normalisatie helpen bij risicoanalyse door inzichtelijk te maken welke factoren bijdragen aan de risico's, hoe groot die rollen zijn, en welke relaties er (mogelijk) bestaan tussen factoren.

We maken onderscheid tussen twee soorten verrijking en normalisatie, op basis van:

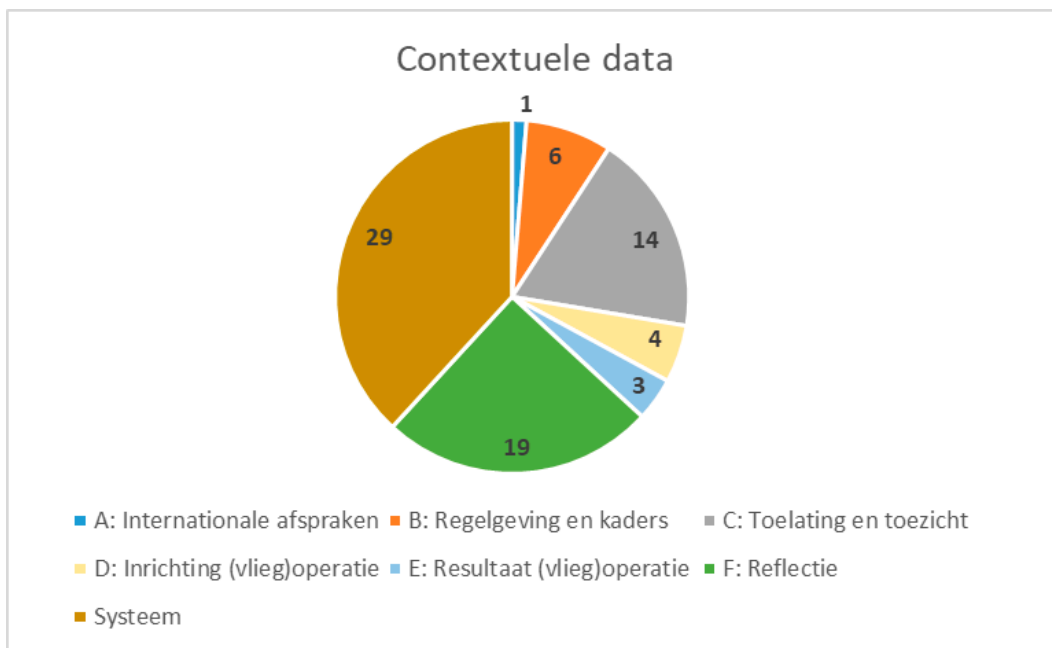
- Activiteitsniveau ('geleverde productie') per systeemelement, bijvoorbeeld het aantal vliegbewegingen, meldingen, uitgevoerde inspecties, uitgevoerde incidentenonderzoeken, verleende vergunningen, inspecteurs, etc.
- Naar omgevingsfactoren of informatie over het 'ecosysteem', bijvoorbeeld weerscondities (wind, neerslag, zicht), runway conditions, vlootsamenstelling, vliegtuiggeneraties, etc.

De dataverrijking en normalisatie die mogelijk en zinvol is, is sterk afhankelijk van het vraagstuk, de gewenste analyse en de beschikbare data. Dit maakt het onmogelijk om op voorhand een complete lijst met mogelijke verrijkingen en normalisaties op te stellen. In dit onderzoek hebben we daarom op basis van beschikbare kennis, ervaring en eerdere veiligheidsanalyses een lijst van mogelijke verrijkingen of normalisaties opgesteld. Per element van het systeemmodel hebben we contextuele data geïnventariseerd die relevant kunnen zijn voor de NVA. Per contextueel datatype is aangegeven of er, en zo ja welke, bronnen aanwezig zijn om de betreffende data te leveren. Hierdoor kunnen we achterhalen voor welke soort verrijking of normalisatie nog geen databronnen beschikbaar zijn of data nog niet wordt verzameld. Indien zinvolle data zijn geïdentificeerd, terwijl een databron nog niet bekend of beschikbaar is voor deze data, dan kan dat aanleiding zijn om te investeren in de vergaring van deze data. Het resultaat van de inventarisatie is vastgelegd in Appendix B. Paragraaf 3.2 geeft enkele voorbeelden van verrijking en normalisatie.

3.2 Resultaten

3.2.1 Overzicht van contextuele data

Er zijn 76 contextuele data geïdentificeerd die kunnen helpen om data te verrijken en te normaliseren (zie Appendix B). Voor 21 (28%) contextuele data zijn (nog) geen databronnen bekend die deze contextuele data zouden kunnen leveren. In Figuur 5 is weergegeven op welk deel van het systeemmodel de contextuele data betrekking hebben of van afkomstig zijn. Bijna de helft van de contextuele data die ingezet kunnen worden voor verrijking en normalisatie heeft betrekking op en is afkomstig uit de (vlieg)operatie.



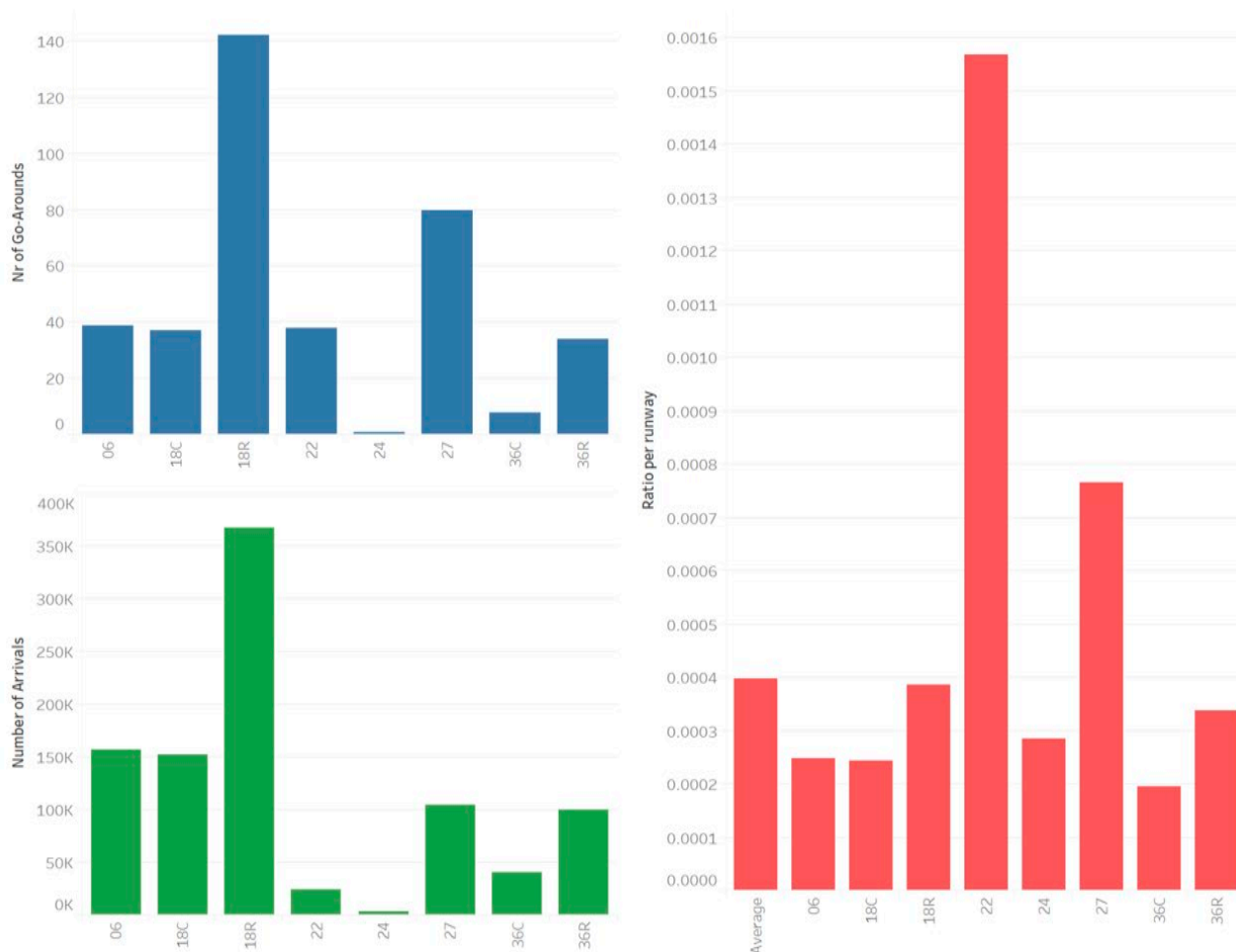
Figuur 5: Het aantal contextuele data per element van het systeemmodel en data met achtergrondinformatie over het luchtvaart 'ecosysteem'.

3.2.2 Voorbeelden van verrijking en normalisatie

In deze paragraaf zijn vijf voorbeelden uitgewerkt met databronnen die in het bezit zijn van het NLR. De figuren zijn bedoeld als illustratie van mogelijke verrijking en normalisatie om te laten zien welke nieuwe inzichten men kan opdoen door het combineren van databronnen. Men kan echter geen conclusies trekken aan de hand van deze voorbeelden omdat de data niet volledig zijn.

3.2.2.1 Go-arounds per landingsbaan en per nadering

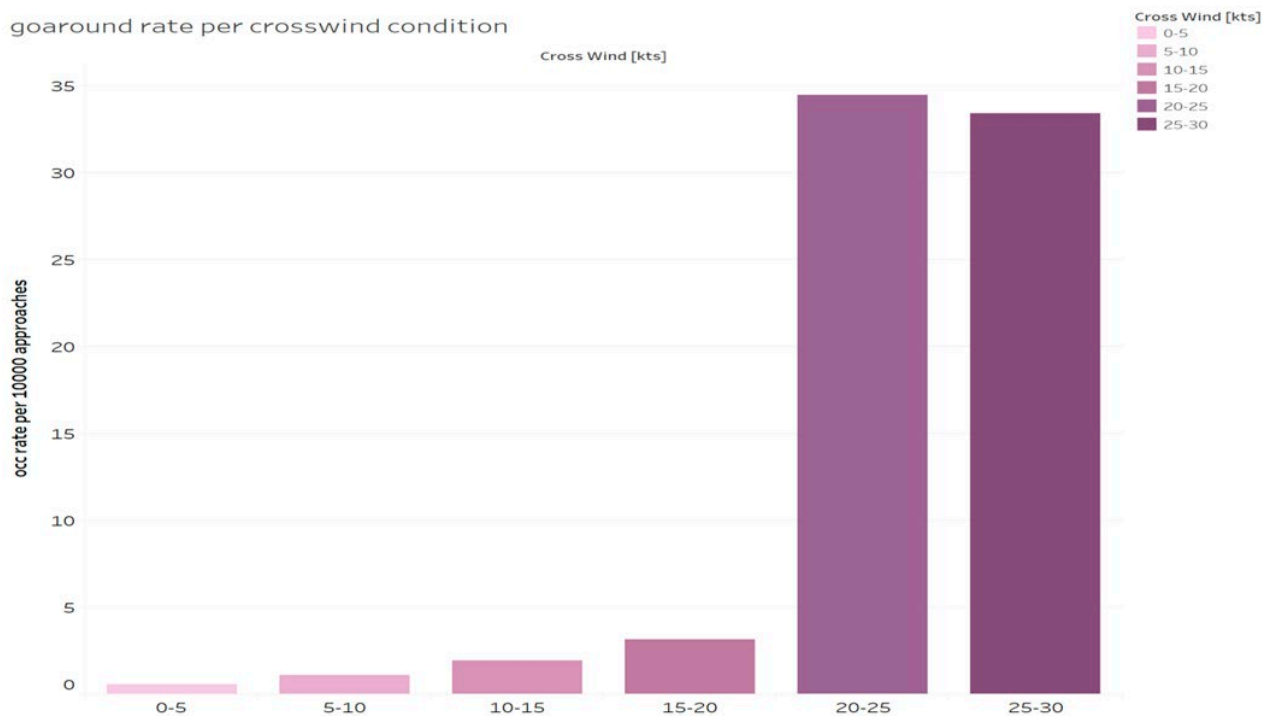
In dit voorbeeld verrijken en normaliseren we het aantal meldingen van go-arounds met de betreffende landingsbaan en het bijbehorende aantal naderingen per jaar per landingsbaan op basis van vluchtdata. Alleen het aantal gemelde go-arounds per jaar geeft namelijk onvoldoende informatie voor verdere analyse. In dit voorbeeld zetten we het aantal go-arounds af tegen het aantal uitgevoerde naderingen op de verschillende landingsbanen. Figuur 6 laat zien dat op baan 18R in aantallen de meeste go-arounds plaatsvinden. Echter, omdat op de ene landingsbaan meer vluchten worden afgehandeld dan op de andere, geeft dit figuur een vertekend beeld. Daarom is ook het aandeel go-arounds ten opzichte van het totaal afgehandelde vluchten (naderingen) per baan berekend. Deze verrijking en normalisatie laat een ander beeld zien. Het aantal go-arounds per nadering op baan 18R blijkt dan gemiddeld te zijn. Daarentegen blijkt het aantal go-arounds op baan 22 per uitgevoerde nadering significant hoger te zijn dan voor de andere landingsbanen. Het aandeel go-arounds per nadering op baan 22 is een factor 5 hoger dan het gemiddelde. Dit voorbeeld toont aan dat verrijking en normalisatie een ander beeld van de situatie kan opleveren.



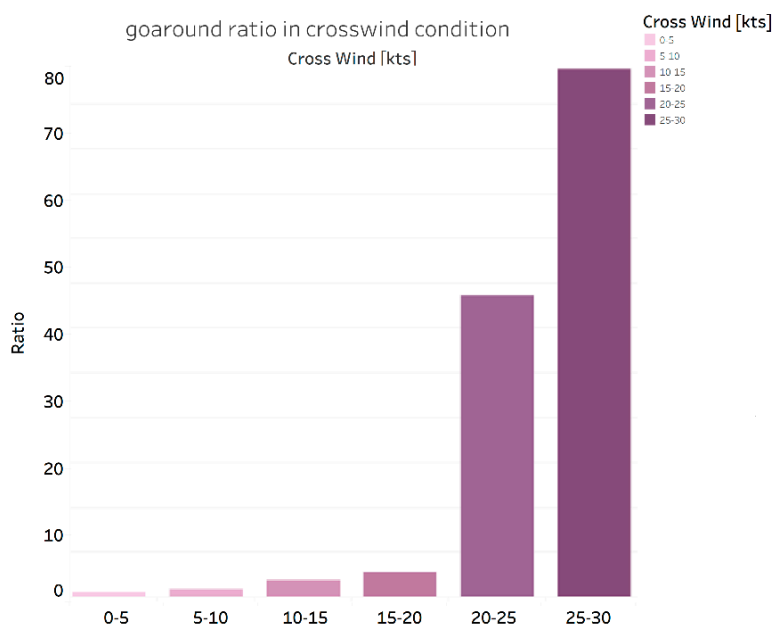
Figuur 6: Aantal meldingen van go-arounds en naderingen per landingsbaan, en de resulterende go-around rate (ratio) per nadering per landingsbaan.

3.2.2.2 Go-arounds in dwarswind

Een tweede voorbeeld van dataverrijking is de integratie van meldingen over go-arounds met weersdata. Het toevoegen van de windconditie aan meldingen van een go-around en het normaliseren naar het aantal uitgevoerde landingen in een bepaalde windconditie levert aanvullend inzicht op ten opzichte van enkel het weergeven van aantallen go-arounds per tijdseenheid. Als eerste stap kan men het aantal go-arounds per aantal landingen bepalen voor verschillende windcondities zoals in Figuur 7 wordt weergegeven. Het figuur laat zien dat het aantal go-arounds per 10.000 naderingen sterk toeneemt bij een dwarswind van meer dan 20 knopen (wat operationeel gezien goed verklaarbaar is). Figuur 8 presenteert het aantal go-arounds per 10.000 naderingen gedeeld door het aantal go-arounds per 10.000 landingen voor de windconditie 0-5 knopen. Hierdoor wordt een ratio berekend waarmee de kans op een go-around kan worden geschat bij een dwarswind X ten opzichte van een go-around zonder dwarswind (dwarswindconditie 0-5 knopen). De figuur laat dus zien dat de kans op een go-around in 20-25 kts dwarswind circa 50x hoger is ten opzichte van de kans op een go-around bij een dwarswindsterkte 0-5 knopen.



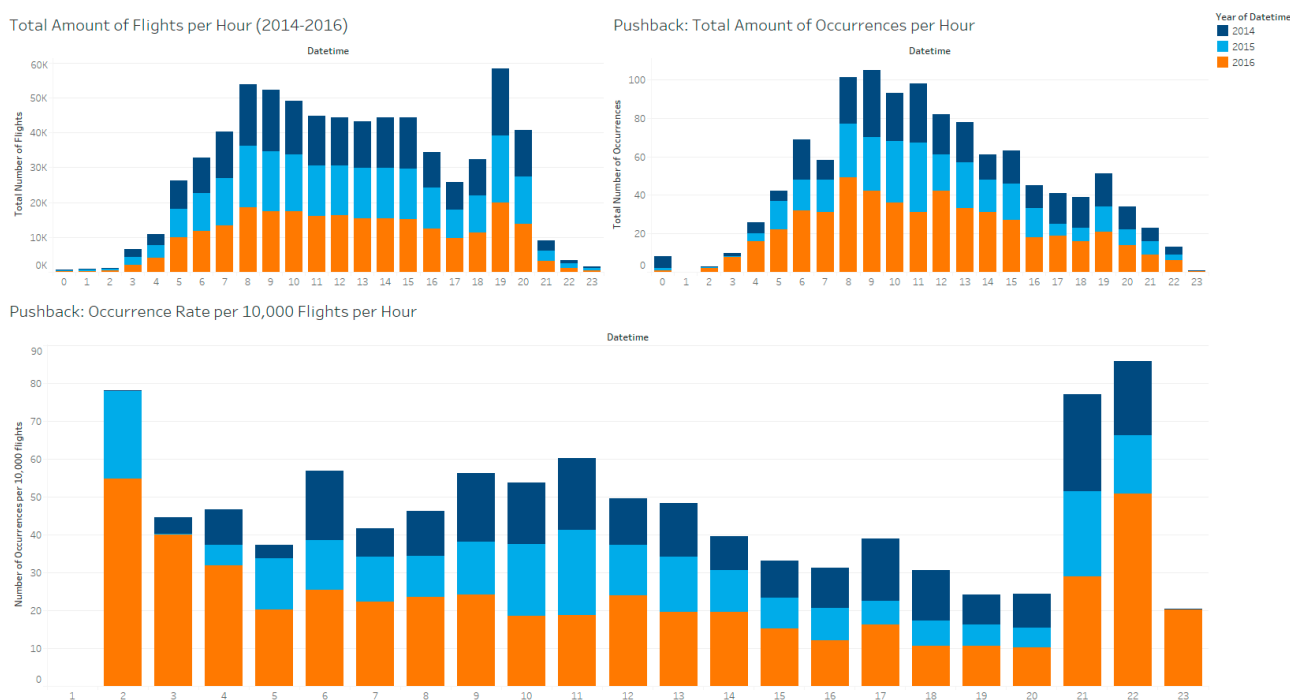
Figuur 7: Aantal go-arounds in een bepaalde dwarswindsterkte ten opzichte van het aantal uitgevoerde landingen in die dwarswindsterkte per 10.000 naderingen (approaches).



Figuur 8: Ratio van de kans op een go-around gegeven een bepaalde dwarswindsterkte ten opzichte van de kans op een go-around gegeven de dwarswindsterkte 0-5 knopen.

3.2.2.3 Voorvallen tijdens pushbacks

In het derde voorbeeld verrijken we het aantal meldingen over voorvallen tijdens de pushback met het tijdstip. Het aantal meldingen van pushbackvoorvallen per jaar levert alleen informatie over de ontwikkeling van de trend op jaarbasis, maar geeft verder geen inzicht in mogelijke oorzaken. Een mogelijke verrijking is om de gemelde pushbackvoorvallen over de uren van de dag inzichtelijk te maken, zoals in Figuur 9 (in het diagram rechtsboven). Dit laat zien dat de meeste voorvallen in de ochtend, tussen 08:00 en 12:00, plaatsvinden. Als men het aantal meldingen normaliseert ten opzichte van het aantal afgehandelde vliegtuigen per uur over de dag, dan toont het onderste staafdiagram in Figuur 9 een ander beeld. Ten eerste zien we een relatief hoger aantal pushback voorvallen in de late avond/nacht. Dit zou kunnen duiden op invloed van lichtcondities en/of vermoeidheid. Daarnaast toont de figuur dat het aantal voorvallen per afhandeling van de vroege ochtend tot de middag redelijk gelijkmatig is en afneemt in de middag. Dit voorbeeld laat zien dat de verrijking en normalisatie van het aantal pushbackvoorvallen, kan helpen bij het identificeren van mogelijke onderzoeksvragen voor een verdiepend onderzoek naar achterliggende factoren.



Figuur 9: Meldingen van pushbackvoorvallen over de dag, en genormaliseerd naar aantal afgehandelde vliegtuigen over de dag.

3.2.2.4 Locatie van runway incursions

In de Staat van Schiphol 2019 [8] staat een verrijking van voorvalldata met de locatie van het voorval: “Het ABL analyseert de locaties in de meldingen van voorvallen die op Schiphol plaatsvinden. De analyses helpen om te leren van voorvallen en brengen mogelijk structurele oorzaken in beeld. De kaart geeft de locaties aan die genoemd worden in meldingen in gebruiksjaar 2019. [...] Locaties in het gebied waar de Buitenveldertbaan en de Aalsmeerbaan elkaar kruisen, zoals N2, komen relatief vaak voor. LVNL identificeert dit gebied ook als hotspot, vanwege een verhoogd risico op taxibaan en runway incursions.” [8]. Door het aantal meldingen en de locatie van het voorval te combineren wordt inzichtelijk welke locaties relatief vaak gemeld worden. Deze informatie kan bijdragen aan het bewustzijn bij

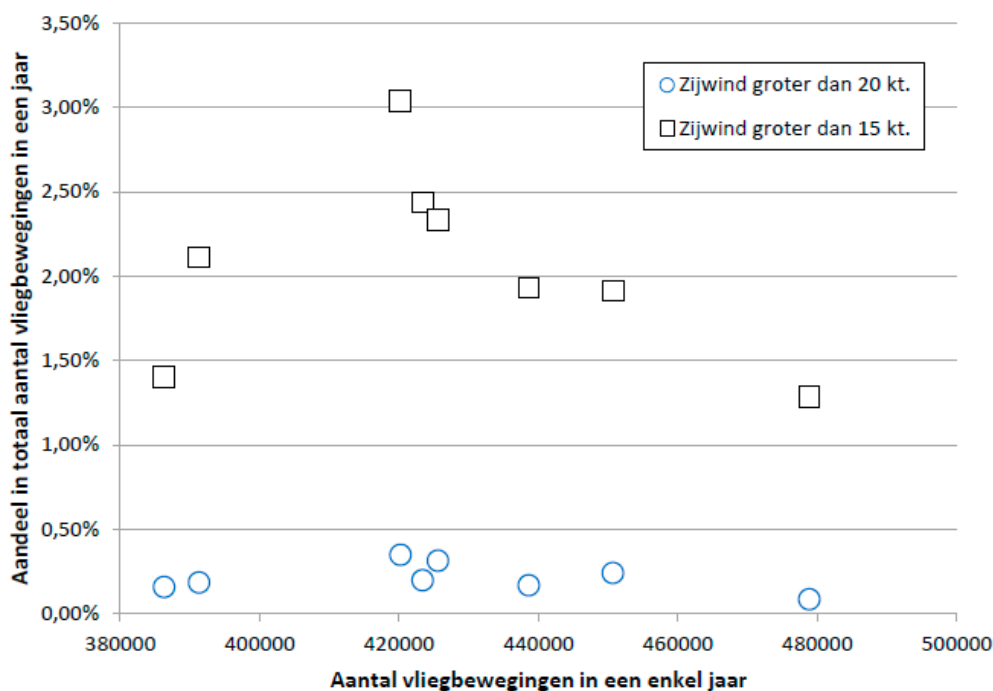
vliegers en verkeersleiders over zogenaamde hotspots. De hoeveelheid verkeer op een bepaalde locatie kan een rol spelen (meer bewegingen op een locatie kunnen bijdragen aan een hoger aantal meldingen over die locatie). Een andere verrijking is het meewegen van het aantal taxibewegingen per locatie.



Figuur 10: Locaties genoemd in voorvalmeldingen en locaties van runway incursions [8].

3.2.2.5 Invloed windcondities en verkeersgroei

Een laatste voorbeeld van dataverrijking komt uit de Integrale Veiligheidsanalyse Schiphol [2]. Beschikbare historische baangebruiksgegevens voor Schiphol gecombineerd met winddata geven inzicht in hoe vaak vliegtuigen onder bepaalde zij- en staartwindcondities starten en landen. Het relatief aantal bewegingen gemaakt onder bepaalde zij- en staartwindcondities in een bepaald jaar, is met behulp van deze data gecorreleerd aan het totaal aantal bewegingen wat in het betreffende jaar gemaakt is. Een deelresultaat van deze correlatie is weergegeven in Figuur 11. Uit deze verrijking blijkt dat er geen invloed is van het aantal jaarlijkse bewegingen op het relatief aantal bewegingen in hoge zijwind. Het relatief aantal starts en landingen in zijwind van meer dan 20 knopen neemt bijvoorbeeld niet toe bij verkeersgroei. Dit aandeel blijft rond 0.2% schommelen zoals blijkt uit deze figuur. Met andere woorden de verkeersgroei in de periode 2009-2016 heeft er niet toe geleid dat er vaker onder hoge zijwind is gestart of geland.



Figuur 11: Aantal vliegbewegingen per jaar uitgezet tegen het relatief aantal operaties in zijwind op Schiphol (2009-2016) [2].

3.3 Beschouwing

Er zijn veel soorten data beschikbaar in databronnen die benut kunnen worden in de analyse van gevaren en risico's, en context kunnen geven in dergelijke analyses. Op hoofdlijnen biedt deze contextuele data informatie over de mate van activiteit of productie van een partij, en/of over het 'systeem' en de omgeving waarin de operatie plaatsvindt. Ter inspiratie staat in Appendix B een lijst met mogelijk contextuele data die gebruikt kan worden voor verrijking en normalisatie.

De voorgaande voorbeelden van dataverrijking en normalisatie laten zien hoe divers verrijking en normalisatie kunnen zijn en welke aanvullende inzichten de combinatie van databronnen kan opleveren. Afhankelijk van het onderwerp en de beschikbare data kan men op verschillende manieren een vraagstuk analyseren. De kennis en ervaring van de analist speelt een belangrijke rol in de keuze van de specifieke verrijking en normalisatie. Echter, in het algemeen kan men stellen dat een normalisatie naar activiteit (aantal vluchten, aantal landingen, aantal inspecties, etc.), naar tijdseenheid (per jaar, kwartaal, maand, etc.), of naar locatie, vluchtfase, baangebruik een goede eerste stap is.

4 Machine Learning

4.1 Inleiding

Een deel van dit onderzoek richtte zich op de toepassing van Machine Learning (ML) ten behoeve van de NVA. Op basis van een literatuurstudie hebben we een aantal interessante toepassingen geïdentificeerd die mogelijk op termijn ingezet kunnen worden om de analyse in de NVA te ondersteunen. De verkenning van Machine Learning technieken en toepassingen omvatte ontwikkelingen bij andere organisaties¹ in de luchtvaart, maar ook toepassingen buiten de luchtvaart die mogelijk bruikbaar of ‘vertaalbaar’ zijn voor de NVA. In dit onderzoek gaat het name om de toepassing van text-mining technieken voor de NVA en veiligheidsanalyse. Deze keuze is gemaakt gelet op de reikwijdte van ML-technieken en mogelijke toepassingen. Het NLR-team heeft enkele Machine Learning technieken op het gebied van text-mining toegepast op drie use cases om de haalbaarheid en bruikbaarheid van deze technieken te evalueren. Paragraaf 4.2 introduceert het onderwerp Machine Learning en beschrijft beknopt de uitkomsten van de literatuurstudie. Appendix C bevat een samenvatting van de literatuurstudie. Paragrafen 4.3 t/m 4.6 beschrijven drie use cases en de resultaten van de toepassing van Machine Learning technieken in deze cases.

4.2 Resultaten literatuurstudie

Machine Learning richt zich op de ontwikkeling, modellering en training van computerprogramma's (algoritmes of modellen) die in staat zijn te leren van (historische) data en vervolgens op basis daarvan een voorspelling kunnen doen. ML-technieken worden in de literatuur op vele manieren geclassificeerd, o.a. op basis van de mate van supervisie. In feite is dit een classificatie op basis van de mate waarin voorkennis met betrekking tot het gedrag van het systeem of de situatie wordt meegenomen in de modelontwikkeling. Een veel gebruikte onderverdeling van technieken is supervised, semi-supervised, reinforcement algorithms en unsupervised:

- Supervised ML-technieken vereisen dat er gelabelde trainingsdata aanwezig is om het model te leren bepaalde input aan bepaalde output te koppelen. De kern van een supervised model is dus het ontwikkelen van een voorspellend model voor bijvoorbeeld het classificeren van data of het voorspellen van bepaalde gedrag op basis van historische data. Een probleem is dat er in de praktijk meestal relatief weinig gelabelde trainingsdata beschikbaar zijn. Een tweede probleem is dat over het algemeen het verkrijgen van gelabelde trainingsdata een omvangrijke en kostbare inspanning is.
- Semi-supervised ML-technieken vereisen alleen dat er gelabelde trainingsdata aanwezig is voor een bepaald 'gedrag'. Methoden voor semi-supervised ML zijn daardoor in het algemeen breder toepasbaar dan supervised ML. Bij supervised machine learning moeten er namelijk trainingsdata zijn voor elke mogelijke uitkomst. Bij semi-supervised machine learning dient er alleen trainingsdata te zijn voor modeluitkomsten die van belang zijn. Een voorbeeld hiervan is het trainen van een model wat nieuwsartikelen in de rubriek sport moet herkennen. Het model dient bij semi-supervised learning alleen getraind te worden op artikelen die daadwerkelijk uit de sport rubriek komen. Het model hoeft hiervoor niet te weten bij welke andere rubrieken de andere artikelen horen. Het probleem van het verkrijgen van gelabelde trainingsdata geldt ook voor semi-supervised technieken, maar in mindere mate dan voor supervised technieken, omdat alleen gelabelde trainingsdata voor normaal gedrag (zoals: de rubriek sport) nodig zijn.

¹ Dit kunnen (inter)nationale instanties, overheden, toezichthouders, en operationele partijen zijn. Bijv. IATA, Eurocontrol, EASA, SESAR, CAAs, sectorpartijen, NASA, MITRE, etc.

- Een reinforcement algorithm is een techniek waarbij het model (neurale netwerk) als het ware wordt beloond voor positieve resultaten en wordt gestraft voor negatieve resultaten. Hierdoor leert het model zichzelf continu te verbeteren. Het betekent wel dat de ontwikkelaar continu de resultaten van het model moet beoordelen. Dit dient ofwel handmatig te gebeuren of met behulp van andere technieken die dit proces automatiseren.
- Unsupervised ML-technieken werken zonder gelabelde trainingsdata. Methoden voor unsupervised ML zijn daardoor in het algemeen breed toepasbaar. Unsupervised technieken zoeken zelf, ongestuurd, naar een logica of patronen in de onderliggende data. Hierbij worden voorspellingen gebaseerd op basis van de samenstelling van de dataset op dat moment. Er wordt geen gebruik gemaakt van historische data.

Op basis van een literatuurstudie zijn een aantal interessante toepassingen en analyses in kaart gebracht, die mogelijk op termijn ingezet kunnen worden om de analyse in de NVA te ondersteunen. De verkenning van Machine Learning technieken en toepassingen omvatte ontwikkelingen bij andere organisaties in de luchtvaart, maar ook toepassingen buiten de luchtvaart die mogelijk bruikbaar of 'vertaalbaar' zijn voor de NVA. Appendix C geeft een samenvatting van de literatuurstudie en gaat in op zes relevante voorbeeldtoepassingen. Deze voorbeelden zijn kort beschreven en geclassificeerd naar onderwerp, categorie en methode van Machine Learning, input en output, en mogelijke toepassing of relevantie met het oog op de NVA.

4.3 Toepassing van ML in drie use cases

Het NLR heeft enkele ML-technieken toegepast in een drietal use cases om de haalbaarheid en bruikbaarheid van deze technieken te evalueren. De onderstaande tabel toont de drie cases. Om te beginnen is het onderwerp en een mogelijke onderzoeksvraag geformuleerd die aan de orde zou kunnen komen in de NVA. Daarna is de categorie en de methode van de ingezette ML-technieken, met input en output, en de toepassing van deze case in het kader van de NVA beschreven.

Tabel 4: Overzicht van de drie use cases.

Case 1	Labelling van occurrences
Onderwerp	<ul style="list-style-type: none"> • Het categoriseren van incidentenrapporten naar 'occurrence category'. Onderzoeksvraag: Welke occurrence category label past bij een bepaald voorval? • Doel is de automatisering van een arbeidsintensief proces en het bereiken van een consequentere manier van labelen. Deze toepassing is interessant voor het ABL of een vergelijkbare partij die voorvalmeldingen categoriseert. • Indien incidentenrapporten gelabeld zijn kan een analist makkelijker onderzoek doen naar specifieke ongevalstypes. Voorbeeld van een analysevraag: Welke voorvallen hadden in potentie tot een runway excursion kunnen leiden?
Categorie	Supervised Machine Learning.
Methodologie	Tekstclassificatie van voorvalrapporten.
Input	Gelabelde CADORS voorvallen database, en ongelabelde ABL voorvallen database.
Output	Gelabelde voorvallen met een CICTT occurrence category label.
Toepassing NVA	Ondersteunt de identificatie van risicofactoren en ondersteunt risicoanalyse. Expertmening is nodig voor het vaststellen van de relevantie en invloed van factoren voor het Nederlandse luchtvaartstelsel.

Case 2	Automatisch genereren van een probable cause op basis van een feitenbeschrijving
Onderwerp	Analyse van oorzaken en verbanden tussen oorzakelijke factoren in een grote verzameling van ongevallen- en incidentenrapporten. Onderzoeksvraag: <ul style="list-style-type: none"> • Welke oorzakelijke factoren komen vaak voor en welke combinatie van oorzakelijke factoren zien we wereldwijd in ongevallen en incidenten? • Welk van deze factoren of patronen zijn relevant voor het Nederlandse luchtvaartstelsel?
Categorie	Supervised Machine Learning.
Methodologie	Text Summariser.
Input	De tekst van NTSB incident- en ongevalsrapporten (factual information/narrative en analysis sections).
Output	Samenvatting van de 'probable cause' per NTSB rapport op basis van de input tekst. Dit is in feite een voorbereidende stap vóór het kunnen analyseren van oorzaken en verbanden tussen oorzakelijke factoren in een grote verzameling van ongevallen- en incidentenrapporten.
Toepassing NVA	Ondersteunt identificatie van risicofactoren en de risicoanalyse. Expertmening is nodig voor het vaststellen van de relevantie en invloed van factoren voor het Nederlandse luchtvaartstelsel.

Case 3	Sentiment analyse op sociale media berichten
Onderwerp	Welke onderwerpen leven er in de sector, bij stakeholders, onder het publiek met betrekking tot de luchtvaart, veiligheid, beleid etc.?
Categorie	Supervised en Unsupervised Machine Learning.
Methodologie	Sentiment Analysis.
Input	Twitter posts (Tweets) geselecteerd op basis van de aanwezigheid van luchtvaart specifieke termen.
Output	Voorspelde sentiment per Tweet.
Toepassing NVA	Ondersteunt identificatie van sentiment over een bepaald onderwerp. Output kan benut worden voor veiligheidspromotie en communicatie.

4.4 Case 1: Automatisch labelen van voorvalrapport met de Occurrence Category

4.4.1 Inleiding

De eerste use case maakt gebruik van een supervised ML-model dat getraind wordt met historische data om een stuk tekst te classificeren. Het ontwikkelde model kan voorvalmeldingen automatisch voorzien van één of meerdere labels van de CICTT occurrence category. Deze taxonomie omvat 36 verschillende voorval categorieën waarmee verschillende soorten voorvallen kunnen worden geclassificeerd. In deze use case is er voor gekozen om voorvalmeldingen te classificeren met de CICTT occurrence category, maar het is ook mogelijk om de data met andere taxonomieën te classificeren. De enige voorwaarde voor de toepassing van het model is dat er voldoende en kwalitatief goede trainingsdata beschikbaar is. Dankzij deze automatische tekstclassificatie kan een arbeidsintensief proces zoals het handmatig toekennen van labels aan voorvalmeldingen worden geautomatiseerd. Tevens kan de tekstclassificatie bijdragen aan een meer uniforme of gestandaardiseerde labelling van data.

4.4.2 Data

Het model is getraind met een dataset van voorvallen uit CADORS (Civil Aviation Daily Occurrence Reporting System) van Transport Canada. Transport Canada is net als ABL verantwoordelijk voor de verzameling en analyse van voorvallen in de luchtvaart. CADORS biedt het publiek de mogelijkheid om voorvalrapporten in te zien en te downloaden. Zodoende zijn er door het NLR 214.177 voorvalrapporten uit CADORS verzameld. Het voordeel van de CADORS dataset is dat deze meldingen door de analisten van Transport Canada zijn gelabeld met (onder andere) de CICTT occurrence category. Ook zijn er, waar toepasbaar, meerdere occurrence category labels per report gelabeld. Dit maakt dat er weinig rapporten met ontbrekende labels zijn in deze dataset, wat het een geschikte training- en testset maakt. Daarnaast bevat elk voorvalrapport onder meer een headline (titel) en een narrative (verslag) waarin het voorval staat beschreven in de vorm van vrije tekst.

4.4.3 Preprocessing

Voordat een ML-model getraind kan worden met behulp van de CADORS dataset moet deze data op verschillende manieren bewerkt worden. Deze stappen zijn er op gericht om de dataset schoon te maken en uiteindelijk in het juiste formaat aan te leveren aan het model. De belangrijkste stappen van dit proces zijn hieronder beschreven.

Preprocessing: data opschonen

Deze stap is erop gericht om de data schoon te maken. Hieronder valt voornamelijk het standaardiseren van de tekst in de rapporten en het verwijderen van stukken tekst die irrelevant zijn voor het bepalen van een occurrence category. Omdat Canada ook Franstalige regio's kent zijn sommige CADORS rapporten in het Frans geschreven. De scope van deze use case was het ontwikkelen van een Engelstalig tekstclassificatie model. Zodoende zijn alle Franstalige rapporten uit de CADORS dataset verwijderd. Hiervoor is gebruik gemaakt van een taaldetectie algoritme dat de voertaal van elk rapport bepaalt. Na deze stap is de tekst in de rapporten schoongemaakt. Hierbij was het van belang dat alle onbruikbare informatie uit de tekst werd verwijderd. Dit is gedaan om zeker te stellen dat het model wordt getraind op relevante informatie. Typische informatie die uit de tekst verwijderd is, is bijvoorbeeld de vliegtuigregistratie, luchthaven ICAO-code en datum. Daarnaast zijn leestekens verwijderd en andere karakters die voor het model geen betekenis aan de tekst toevoegen. Deze schoonmaakstappen zijn gedaan met regular expressions in de Python Programming Language, waarmee de hierboven genoemde stukken tekst en leestekens zijn verwijderd. Na het opschonen van de dataset is 80% van de dataset gebruikt om het model te trainen en is 20% van de dataset gebruikt om het model te testen.

Bij het gebruik van klassieke tekst classificatie algoritmes zijn nog andere preprocessing stappen nodig, zoals het verwijderen van stopwoorden en het vervangen van synoniemen voor een enkele term. Deze stappen bleken niet nodig bij het gebruik van het algoritme in deze use case.

Preprocessing: data converteren

Waar bij andere klassieke modellen de tekst op bepaalde manieren geconverteerd dient te worden naar een vector, is dat bij het gebruik van het BERT-model overbodig. Het model heeft eigen ingebouwde methodes die dit voor de gebruiker doen.

4.4.4 Model

Voor het classificatie algoritme is gebruik gemaakt van het BERT-model [29]. Het BERT-model is een neurale netwerk dat gebruik maakt van een transformer architectuur. Het model gebruikt transfer learning waarbij de training bestaat uit een pre-training fase en een fine-tuning fase. Tijdens de pre-training fase wordt het model getraind op een triviale taak waar veel data voor beschikbaar is. In het geval van BERT zijn twee pre-training taken gebruikt. De eerste taak is de zogenaamde 'masked language model' (MLM). Hierbij worden willekeurige woorden in de input tekst vervangen door een mask token. Het is dan de taak van het algoritme om te voorspellen welk woord oorspronkelijk stond op de plaats van het mask token. De tweede pre-training taak is 'next sentence prediction' (NSP). Tijdens NSP worden twee zinnen als input gegeven aan het model. Het model wordt dan getraind om te voorspellen in welke volgorde deze twee teksten zouden moeten staan. Voor de pre-training is gebruik gemaakt van BooksCorpus en wikipedia data. Tijdens de fine-tuning fase wordt een laag toegevoegd aan het model, waarna het model getraind wordt op de specifieke taak waarvoor het toegepast zal gaan worden. In het model dat hier ontwikkeld is bestaat deze toegevoegde laag uit een logistisch regressie model, waarvan de input het CLS token uit het BERT-model is. De transfer learning benadering van het BERT-model heeft als beoogd voordeel om het gebruik van neurale netwerken, die hoge prestaties behalen maar veel trainingsdata vereisen, ook te kunnen gebruiken in situaties waar relatief weinig data beschikbaar is.

4.4.5 Resultaat

Na het trainen van het model op de CADORS dataset is het model geëvalueerd met de testdata van CADORS. Daar werd een betrouwbaarheid van omstreeks 90% per label behaald. Hierna werd het model toegepast op een andere dataset met voorvalmeldingen uit de jaren 2014 t/m 2016 die nog niet gelabeld waren. Aan de 42.188 rapporten in deze nieuwe dataset zijn één of meerdere occurrence category labels toegekend door het model. Tabel 5 toont als voorbeeld het resultaat van de toegekende labels aan enkele rapporten.

Tabel 5: Voorbeelden van voorvalmeldingen gelabeld door het model.

Report	Label
Hard landing. Hard landing upon arrival. First on main gear, then also on nose gear. The aircraft did not bounce during touchdown. [...]. Requested hard landing inspection.	['Abnormal runway contact']
Hard landing report. Aircraft produced hard landing report after landing. During approach at [airport], the aircraft was flown stable in gusty conditions. Due gusts corrections were needed, resulting in a firm landing. The firm landing was considered acceptable by crew.	['Abnormal runway contact', 'Turbulence encounter']
Arriving a/c reports a motorglider at same altitude during final approach Arriving [aircraft] reports a motorglider at same altitude (2000ft) during final approach. Glider was not visible on radar.	['Airprox/TCAS alert/loss of separation/(near) midair collisions']
Wake encounter. Preceding aircraft B744 5 NM ahead caused wake during landing phase at 100 feet. Bank was appr 5 degrees, considerable amount of aileron input needed to compensate for bank. Aircraft touched down with no bank and normal rate and pitch. Crosswind of 7 kts, stable atmospheric circumstances.	['Turbulence encounter']

Prestatie van het model

Een veiligheidsanalist heeft in een niet-random steekproef onder een aantal CICTT occurrence category labels de door het model toegekende labels handmatig geëvalueerd. Deze labels zijn geselecteerd omdat uit ervaring is gebleken dat zij vaak worden gebruikt, een specifieke type occurrence beschrijven met weinig kruisverwijzingen naar andere labels en omdat zij beschikken over een eenduidige definitie wat het handmatig evalueren sneller maakt. Daarbij is vastgesteld of het label juist, gedeeltelijk juist of onjuist was. Een gedeeltelijk juist label kan bijvoorbeeld zijn vastgesteld in een situatie waarin meerdere labels zijn toegekend aan een voorvalmelding, waaronder een juist en een onjuist label. Tabel 6 toont de resultaten van de handmatige evaluatie en laat zien dat de prestatie van het model uiteenloopt in de verschillende occurrence categorieën. In sommige categorieën zoals ARC en BIRD levert het model goede prestaties. In bijvoorbeeld de categorie LOC-G scoort het model daarentegen matig. Wat de oorzaak hiervan is moet verder worden onderzocht.

Tabel 6: Evaluatie van toegekende labels in een beperkte steekproef (aantallen en afgerond percentage).

Label	Foutief (%)	Gedeeltelijk juist (%)	Juist (%)	Totaal (gedeeltelijk juist) (%)	Totaal aantal meldingen
ARC	18 (10%)	20 (12%)	135 (78%)	155 (90%)	173
RE	5 (33%)	7 (47%)	3 (20%)	10 (67%)	15
ICE	21 (30%)	2 (3%)	48 (68%)	50 (70%)	71
LOC-I	3 (10%)	9 (31%)	17 (59%)	26 (90%)	29
LOC-G	36 (58%)	8 (13%)	18 (29%)	26 (42%)	62
BIRD	13 (0%)	58 (2%)	3736 (98%)	3794 (100%)	3807

Runway Excursion analyse

Als incidentenrapporten gelabeld zijn kan een analist makkelijker onderzoek doen naar specifieke ongevalstypes doordat relevant rapporten dankzij de labels eenvoudig op te vragen zijn. Stel dat de analist geïnteresseerd is in het onderwerp 'runway excursion' (RE), dan kunnen dankzij de uitgevoerde labelling voorvallen over runway excursions snel gevonden worden in de database.

Het label 'RE' wordt formeel toegekend aan voorvallen die daadwerkelijk een runway excursion zijn. Deze voorvallen zijn echter zeldzaam in de dataset. Het blijkt dat een aantal voorvallen het label RE heeft gekregen, hoewel er geen echte runway excursion is opgetreden. Bij nadere bestudering blijkt dat een deel van deze meldingen zogenaamde precursors van runway excursions bevatten. Het model labelt dus ook voorvallen waarin precursors van runway excursions zijn opgetreden als 'RE', aangezien de logica van het model deze precursors herkend uit een trainingset met runway excursion voorvallen. Specifieke woorden of woordcombinaties in voorvalmeldingen van daadwerkelijke runway excursions kunnen ook voorkomen in voorvallen waarin geen runway excursion is opgetreden. De analist kan dus ook voorvalmeldingen die in potentie hadden kunnen eindigen in een runway excursion opzoeken onder het label 'RE'. In een eerste stap kan de analist bijvoorbeeld zoeken naar voorvallen met het label ARC (abnormal runway contact) en RE (runway excursion) om een inventarisatie te doen van voorvallen die mogelijk gerelateerd zijn aan runway excursion. In een tweede stap zal de analist deze lijst met voorvallen handmatig moeten doornemen om vast te stellen of er wel/niet sprake is van een runway excursion, en welke mogelijke precursors van runway excursions zijn opgetreden.

4.4.6 Conclusie

Het ontwikkelde model kan voorvalmeldingen automatisch voorzien van één of meerdere labels van de CICTT occurrence category. In dit onderzoek volgt uit een beperkte, niet-random, steekproef (4157 meldingen) dat het model gevarieerd presteert in het correct labelen van voorvalmeldingen. Voor bepaalde type meldingen (zoals bird strikes) levert het model een hoge mate van correcte labels op, terwijl voor sommige type meldingen de prestatie van model matig is. Het nadeel was dat we in dit onderzoek niet de beschikking hadden over een dataset met gelabelde voorvallen waarop dit model automatisch geëvalueerd kan worden (gelabelde data anders dan de CADORS set waarmee het model getraind is). Het gevolg is dat veiligheidsanalisten de toegekende labels handmatig moeten evalueren, wat een arbeidsintensief proces is. Een uitgebreidere validatie van de resultaten is nodig om vast te stellen wat de correctheid van de labelling per occurrence category is. Uit nader onderzoek moet blijken wat de reden is dat het model gevarieerd presteert op verschillende categorieën.

4.4.7 Aanbevelingen

Het is noodzakelijk om de resultaten verder te valideren met een veiligheidsanalist. Daarnaast wordt aanbevolen om onderzoek te doen naar de mogelijke oorzaken van de variërende prestatie van het model op verschillende occurrence categorieën die in de steekproef zijn geëvalueerd.

4.5 Case 2: Automatisch genereren van een probable cause op basis van een feitenbeschrijving

4.5.1 Inleiding

Deze paragraaf beschrijft hoe een Machine Learning model getraind en toegepast is om een tekst te genereren waarin de oorzaak van een ongeval wordt samengevat in 1 tot 2 zinnen aan de hand van een ongevals- of incidentenrapport. Deze taak is vergelijkbaar met het genereren van een samenvatting. Het genereren van samenvattingen aan de hand van algoritmes is een taak die veel onderzocht is in de natuurlijke taalverwerking. In dit onderzoek wordt gebruik gemaakt van het BART-large model [30] dat wordt gefinetuned aan de hand van data van de National Transportation Safety Board (NTSB). Dit model zou onderdeel kunnen zijn van een proces waarbij in combinatie met andere Machine Learning technieken meer inzicht verschaft kan worden in welke oorzaken onderliggend zijn aan ongevallen en incidenten aan de hand van een grote dataset van rapporten.

4.5.2 Data

Om het BART-large model te trainen wordt gebruik gemaakt van de data van de NTSB. De NTSB is een Amerikaans agentschap dat onderzoek doet naar ongevallen en incidenten. De ongevals- en incidentenrapporten van de NTSB kunnen worden verkregen via de NTSB aviation accident database. Deze database bevat rapporten over ongelukken en incidenten die vanaf 1962 hebben plaatsgevonden. De voorvallen die hierin beschreven staan betreffen

grotendeels ongevallen en incidenten uit de civiele luchtvaart, waaronder commercieel luchtverkeer, kleine luchtvaart en zakenverkeer. De rapporten beslaan vluchten die plaats hebben gevonden binnen de gebieden en eigendommen van de Verenigde Staten en internationale wateren ([26], [28]). De rapporten van de NTSB volgen over het algemeen een vaste structuur. Naast gestructureerde data bevatten de rapporten een aantal paragrafen met vrije tekst. Deze paragrafen bestaan uit 'analysis', 'factual information' en 'probable cause and findings'. Het 'factual information' stuk bevat feitelijke informatie en een algemene beschrijving van de vlucht. Het 'analysis' stuk bevat informatie en een analyse van de feiten die direct relevant zijn voor het voorval. In de 'probable cause and findings' wordt uiteindelijk in een of twee zinnen beschreven wat de oorzaak is van het voorval en welke factoren daaraan hebben bijgedragen.

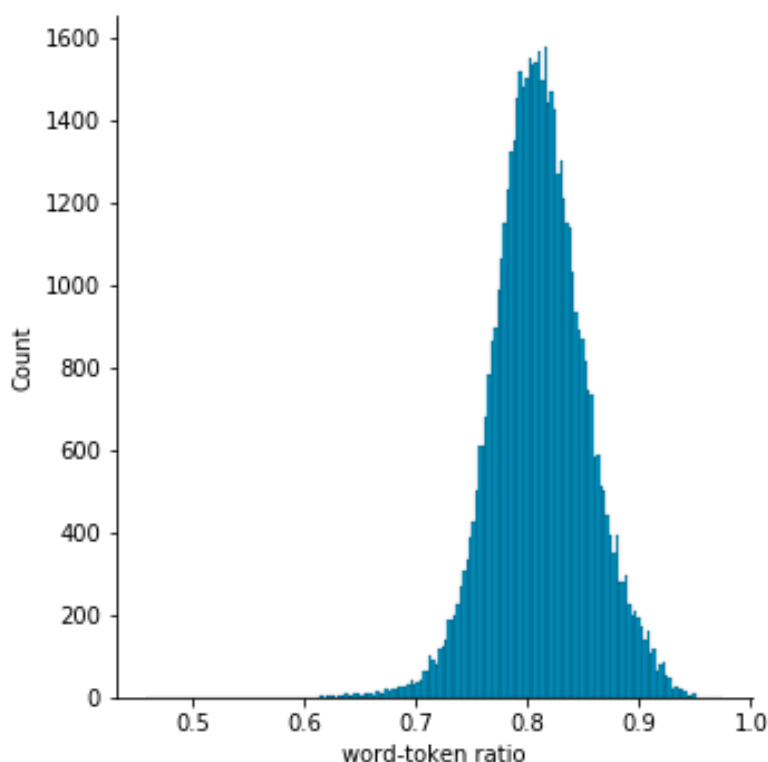
4.5.3 Tokenizer

Om met Machine Learning modellen tekstuele data te kunnen verwerken moet de tekst eerst omgezet worden naar numerieke representaties. Deze numerieke representaties worden ook wel embeddings genoemd. In deze embeddings ligt de betekenis van de tokens gecodeerd. Bij het BART-model bestaan de tokens uit woorden of delen van woorden. In Tabel 7 staat als voorbeeld een 'probable cause' sectie van een rapport uitgedrukt in tekst en als een reeks van tokens. Elk woord wordt omgezet in een of meerdere tokens. Het woord 'loss' wordt bijvoorbeeld omgezet in de token 'Ġloss'. Andere woorden worden omgezet in meerdere tokens. Een voorbeeld daarvan is het woord 'inflight', dat bestaat uit de tokens 'Ġinf' en 'light'. De tokenizer maakt onderscheid tussen tokens die afkomstig zijn van een geheel woord of het begin van een woord, en tokens die afkomstig zijn van een later deel van een woord. Ook worden tokens toegekend aan leestekens. Tenslotte is te zien dat de tokenizer gevoelig is voor hoofdletters.

Tabel 7: Voorbeeld van een 'probable cause' stuk uit een ongevalsrapport en dezelfde tekst uitgedrukt in tokens. De tekst in het rapport bestaat uit 28 woorden. Dezelfde tekst uitgedrukt in tokens bestaat uit 37 tokens.

Probable cause sectie	Tokens
An inflight loss of control on landing approach due to encountering a down-draft, and the pilot's failure to take timely action to abandon the approach and perform a go-around.	'An', 'Ġinf', 'light', 'Ġloss', 'Ġof', 'Ġcontrol', 'Ġon', 'Ġlanding', 'Ġapproach', 'Ġdue', 'Ġto', 'Ġencountering', 'Ġa', 'Ġdown', '-', 'draft', ',', 'Ġand', 'Ġthe', 'Ġpilot', '"', 's', 'Ġfailure', 'Ġto', 'Ġtake', 'Ġtimely', 'Ġaction', 'Ġto', 'Ġabandon', 'Ġthe', 'Ġapproach', 'Ġand', 'Ġperform', 'Ġa', 'Ġgo', '-', 'around', '.'

De maximale lengte van een input tekst is gedefinieerd in het aantal tokens. Een tekst bestaat uit meer tokens dan uit afzonderlijke woorden omdat woorden uit een of meerdere tokens bestaan en omdat er ook tokens gedefinieerd zijn voor leestekens. Het aantal tokens wordt omgezet naar het aantal woorden, aangezien dit een praktischere eenheid is. In Figuur 12 is de verdeling te zien van het aantal input woorden gedeeld door het aantal tokens. Hieruit kan worden geconcludeerd dat deze verhouding ongeveer gelijk is aan 0.8.



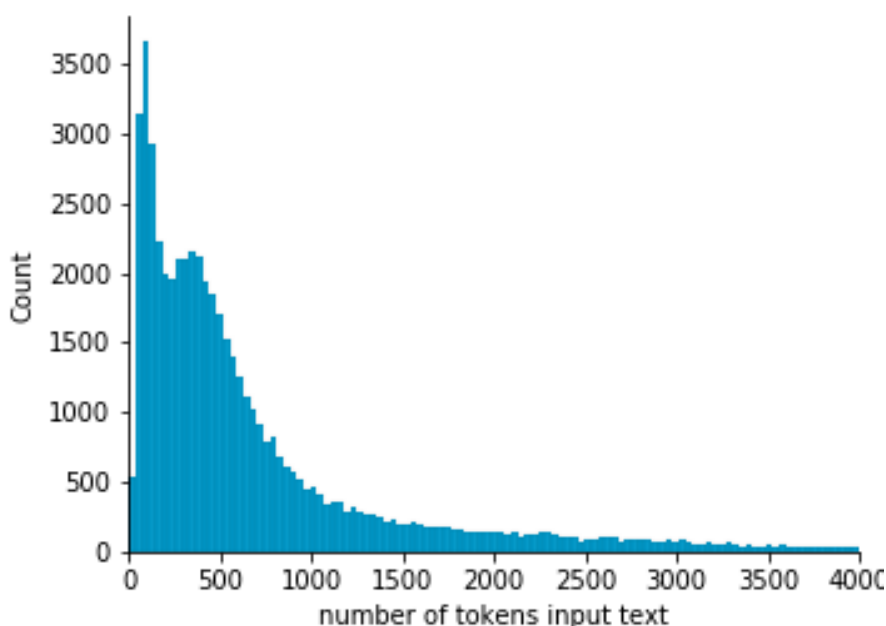
Figuur 12: Verdeling van de woord-token ratio van de input teksten.

Modellen met een transformer architectuur zoals het BART-model vereisen een grote hoeveelheid geheugen. In dit onderzoek is gebruik gemaakt van een GPU met 32 GB geheugen. Voor deze hardware is gebleken dat voor het finetunen van het BART-large model er een praktische limiet is van 1024 tokens. Bij een woord-token ratio van 0.8 betekent dat dat de maximale lengte van de input tekst op ongeveer 800 woorden uitkomt.

4.5.4 Pre-processing

De complete dataset bestaat uit 86.005 rapporten. Voordat de data gebruikt kan worden voor training van het BART-large model worden een aantal data cleaning (opschoning) stappen toegepast. Rapporten waarin de ‘probable cause’ sectie ontbreekt en waarin de ‘factual information’ en ‘analysis’ allebei ontbreken worden verwijderd. Verder worden rapporten die bestaan uit duplicaten verwijderd. Na het uitvoeren van deze stappen bestaat de dataset uit 54.411 rapporten. Een deel van de data bestaat volledig uit hoofdletters. Omdat de tokenizer en het model gevoelig zijn voor hoofdletters worden alle teksten die volledig uit hoofdletters bestaan omgeschreven naar kleine letters. De ‘factual information’ en ‘analysis’ secties in de rapporten worden in deze volgorde samengevoegd tot één input tekst. De keuze voor deze volgorde is willekeurig. De ‘probable cause and findings’ wordt gebruikt als target tekst of referentietekst.

Figuur 13 laat de verdeling van het aantal tokens van de input tekst zien voor de gehele collectie rapporten. Bij het gebruikte model is de maximaal mogelijke lengte van de input tekst 1024 tokens, wat overeenkomt met circa 800 woorden. Daarom wordt alleen gebruik gemaakt van rapporten waarbij de input tekst korter is dan 1024 tokens. Na het toepassen van deze beperking blijven er 42.466 rapporten over, wat neerkomt op 78% van het totaal aantal rapporten.



Figuur 13: Verdeling van het aantal tokens in de input tekst.

De opgeschoonde dataset wordt vervolgens verdeeld volgens de verhouding 70%/15%/15% in respectievelijk een training-, evaluatie- en testset. Hierna bestaat de trainingset uit 29.726 rapporten en de evaluatie- en testset beide uit 6.370 rapporten. De trainingset wordt gebruikt bij het trainen/finetunen van het model. De evaluatieset kan worden gebruikt voor hyperparameter tuning. De testset wordt gebruikt om na afloop een objectieve maat te geven van de prestaties van het model.

4.5.5 Model

Het model dat gebruikt is in dit onderzoek is het BART-large model [30]. Bij de implementatie is gebruik gemaakt van de simple transformers module. Dit is een wrapper van de transformers bibliotheek [34]. Het BART-model maakt gebruik van een encoder-decoder architectuur [30]. Het model kan gezien worden als een generalisatie van het BERT [29] en het GPT [32] model omdat de encoder 'bi-directional' transformer is en de decoder is een 'link-naar-rechts' decoder.

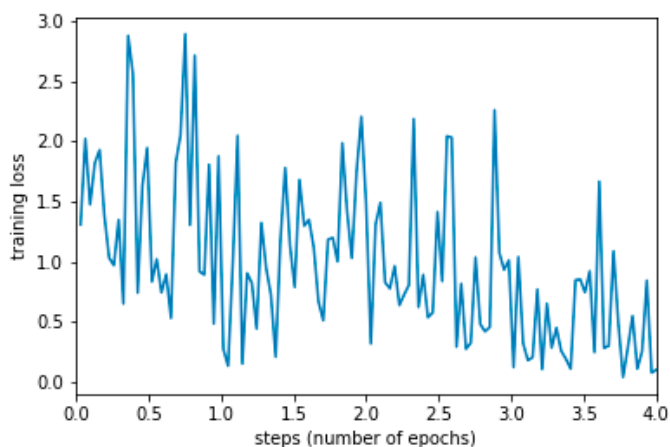
Het proces waarbij een model eerst wordt getraind op een andere taak dan waarop dit model uiteindelijk toegepast gaat worden staat bekend als transfer learning. Tijdens het pre-training proces doet het model generieke kennis op die ook van pas kan komen bij het uitvoeren van de uiteindelijke taak. Het verder trainen van het model op de taak waarvoor het uiteindelijk bedoeld is, wordt ook wel finetunen genoemd. Het BART-model is een model dat gebruik maakt van transfer learning, waarbij de pre-training fase is uitgevoerd door onderzoekers van Facebook en vervolgens publiek beschikbaar is gemaakt.

Het model is gepre-traind op twee taken. Bij de eerste taak worden de zinnen van de input tekst in een willekeurige volgorde gezet. De taak van het algoritme is dan om de tekst in de oorspronkelijke volgorde te reproduceren. In de tweede taak wordt een willekeurig aantal tokens vervangen door het mask token. In dat geval is het de taak van het algoritme om te voorspellen wat de oorspronkelijke tekst was op de locatie van het mask token [30].

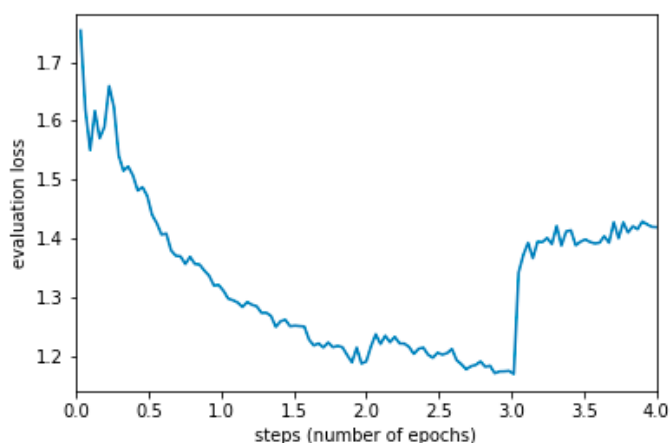
Tijdens het finetunen is gekozen voor een maximale input lengte van 1024 tokens. Daarbij is een training batch grootte en evaluation batch grootte van beide 1 gebruikt. Een grotere batch was niet mogelijk op de beschikbare hardware. Er is gebruik gemaakt van een AdamW optimizer met een learning rate van $\alpha = 5 \cdot 10^{-5}$. Voor de decoder is gebruik gemaakt van een combinatie van top-k en top-p sampling, met als waardes respectievelijk 50 en 0.95.

Het model wordt tijdens het finetunen getraind voor 4 epochs. In 1 epoch wordt het model één keer getraind op alle voorbeelden in de trainingset. De trainingset bestaat uit 29.726 rapporten en de batch grootte is 1, dus in totaal voert het model in 4 epochs $4 \times 29.726 = 118.904$ stappen uit. Tijdens elk van deze stappen voert de optimizer een update van de parameters van het model uit. Na elke 1000 stappen is de loss berekend in zowel de training- als de evaluatieset. Deze waardes zijn te zien in Figuur 14 en Figuur 15.

Beide figuren laten zien dat de loss daalt als functie van het aantal stappen. Het sterk fluctueren van de training loss wordt veroorzaakt door het feit dat de training loss wordt berekend over de training batch, die in deze studie een grootte van 1 heeft. De evaluatie loss daarentegen wordt berekend over de gehele evaluatie set en zodoende gemiddeld. Ook is te zien dat de evaluatie loss toe neemt na 3 epochs terwijl de training loss blijft dalen. Dit is een indicatie van overfitting. Daarom wordt gebruik gemaakt van early stopping. De parameterwaardes die gebruikt worden om voorspellingen te doen zijn in principe de waardes van de parameters waarbij de evaluatie loss minimaal is. In Figuur 15 is te zien dat dit punt precies aan het eind van de derde epoch ligt.



Figuur 14: Training loss als functie van het aantal stappen in epochs.



Figuur 15: Learning curve van de evaluatie loss. Het aantal training stappen is uitgedrukt in epochs.

4.5.6 Resultaten

Hieronder staan een aantal voorbeelden van de resultaten van de toepassing van het getrainde model op de validatieset met ongevals- en incidentenrapporten. Tabel 8 bevat twaalf oorspronkelijke 'probable cause' secties uit NTSB rapporten en daarnaast de gegenereerde samenvatting, d.w.z. de voorspelde samenvatting van de probable cause door het model. Tabel 9, Tabel 10, en Tabel 11 bevatten drie voorbeelden van een input tekst, een probable cause sectie en een gegenereerde samenvatting. Deze rapporten zijn afkomstig uit de evaluatieset. In de onderstaande tabellen is te zien dat het model grammaticaal correcte tekst genereert en dat deze teksten ook inhoudelijk veelbelovend zijn met betrekking tot de juistheid en kwaliteit van de samenvatting. Om een kwalitatieve inschatting te kunnen maken van de prestaties van het model is een systematische evaluatie door een expert nodig. In de sectie met aanbevelingen is hierover meer te lezen.

Tabel 8: Twaalf willekeurig geselecteerde probable cause stukken uit rapporten en de samenvattingen die gegenereerd zijn door het model aan de hand van de input teksten.

Probable cause sectie NTSB rapport	Gegenereerde tekst (samenvatting probable cause)
The pilots failure to maintain directional control during landing roll that resulted in a collision with an airport sign. Factors were wet runway surface and wind gust.	The pilot's failure to maintain directional control during landing. A factor was the wind gust.
Failure of the flight control lower mixing unit due to fatigue, which resulted in the loss of control.	A fatigue failure of the lower mixing unit. contributing to the accident was unsuitable terrain.
A loss of left braking action for reasons that could not be determined because postaccident examination did not reveal any evidence of preimpact mechanical failures or malfunctions that would have precluded normal operation, which resulted in a loss of directional control.	The pilot's inability to maintain directional control during the landing roll due to a lack of left brake action for reasons that could not be determined.
The pilot's failure to maintain directional control during the landing roll, which resulted in a ground loop.	The pilot's failure to maintain directional control during the landing roll.
failure of the number three exhaust valve on the left engine. a major factor was the high density altitude where the airplane would not operate on single engine.	The failure of the number three exhaust valve. a factor in the accident was the high density altitude.
The pilot's failure to maintain directional control of the airplane while landing, which resulted in a runway excursion.	The pilot's failure to maintain directional control during the landing.
The pilot-in-command did not maintain control of the airplane and did not perform the appropriate procedures for an aborted takeoff. a factor in the incident was the asymmetry in power between the plane's two engines for undetermined reasons.	The delayed or decreased torque on the no. 2 engine for undetermined reasons.
An undetermined event which resulted in engine failure on landing roll out during attempted restart fuel ignited and an undetermined event in the airplane's fuel system which resulted in a fire.	The loss of engine power for undetermined reasons.
The pilot's failure to maintain directional control. A related factor was the snow-covered runway conditions.	The pilot's failure to maintain directional control during the takeoff roll. A factor was the snow covered runway.
The pilot's misjudged distance/altitude during landing, which resulted in an undershoot and on ground collision with a dirt bank.	The pilot's failure to maintain a proper glide path during the final approach, resulting in an undershoot and subsequent collapse of the main landing gear.

Probable cause sectie NTSB rapport	Gegenereerde tekst (samenvatting probable cause)
The student pilot's failure to maintain directional control of the aircraft during the landing roll. contributing to the accident was the student's lack of total experience.	Directional control was not maintained by the student pilot.
The pilot's improper recovery from a bounced landing and his failure to maintain control of the airplane during landing.	The pilot's failure to maintain directional control during landing. A contributing factor was the high density altitude.

Tabel 9: Voorbeeld van de input tekst, probable cause sectie en gegenereerde tekst die horen bij een ongevalsrapport.

<p>Input tekst: On January 13, 2002, at 0915 eastern standard time, a Cessna 172M, N9334H, operated by a private pilot, under the provisions of Title 14 CFR part 91, experienced a wind gust on landing. The personnel flight at Sanford Lee County Airport, Sanford, North Carolina. Visual meteorological conditions prevailed at the time of the accident and no flight plan was filed. The private pilot and his passengers did not received any injuries. The airplane sustained substantial damage. The flight departed Sanford Lee County Airport, in Sanford, North Carolina, at 0915. According to the pilot, he experience a strong crosswind from the right while on approach for runway 21. The pilot stated that he was crabbed about 25 degrees toward the runway, at 85 knots abeam the numbers. The pilot stated that when he turned final he added his third notch of flaps and trimmed the airplane to 70 knots. The pilot stated when he was certain he could make the field he added the last notch of flaps. The pilot stated that when the airplane touched down, the nose of the airplane went to the right. The airplane then started a right skid. The pilot stated that the airplane "tipped to the left" and he lost directional control. The airplane skidded off of the right side of the runway, and collided with a runway sign. Examination of the airplane revealed, the left main landing gear collapsed. The left stabilizer was damaged and a hole was punctured in the right stabilizer. The examination of the airplane failed to reveal a mechanical problem. According to the pilot, the runway surface was wet from rain the previous night. The reported winds at the time of the accident were 290 degrees at 10 knots with gust to 18 knots. The pilot reported experiencing a wind gust on landing. The pilot stated that when the airplane touched down, the nose of the airplane went to the right. The airplane then started to skid to right. The airplane continued off the runway and collided with an airport sign. Examination of the airplane failed to disclose a mechanical malfunction.</p>
<p>Probable cause sectie: The pilots failure to maintain directional control during landing roll that resulted in a collision with an airport sign. Factors were wet runway surface and wind gust.</p>
<p>Gegenereerde samenvatting: The pilot's failure to maintain directional control during landing. A factor was the wind gust.</p>

Tabel 10: Voorbeeld van de input tekst, probable cause sectie en gegenereerde tekst die horen bij een ongevalsrapport.

<p>Input tekst: On June 11, 2015, about 0840 mountain standard time, a Beech E33A, N8401N, exited the runway landing surface during the landing rollout at La Cholla Airpark, Tucson, Arizona, and collided with a drainage ditch. The owner/pilot was operating the airplane under the provisions of 14 Code of Federal Regulations Part 91. The commercial pilot was not injured; the airplane sustained substantial damage. The local personal flight departed Tucson, about 0800. Visual meteorological conditions prevailed, and no flight plan had been filed. In a written statement, the pilot reported that the landing touchdown was smooth, with power nearly off. During the landing roll, he applied gentle pressure to the brakes to slow down, however the left side braking action was non-existent, while the right side brake held. The pilot was unable to maintain directional control as the airplane departed the narrow runway and subsequently collided with a drainage ditch and low cement wall. The wreckage was recovered to a secured facility where a Federal Aviation Administration Inspector from the Scottsdale, Arizona, Flight Standards District Office examined the brake system. The inspector reported no evidence of a mechanical malfunction or failure. The hydraulic system was almost empty due to the bleeder valve on the brake caliper was sheared off. The airplane touched down smoothly with the power nearly off. During the landing roll, the commercial pilot applied gentle pressure to the brakes to slow down; however, there was no left braking action, but there was right braking action. The pilot was unable to maintain directional control of the airplane, and it departed the narrow runway and subsequently collided with a drainage ditch and cement wall. During postaccident examination of the brake system, the hydraulic system was found almost empty because the bleeder valve on the brake caliper had sheared off due to impact damage. The examination revealed no evidence of any preimpact mechanical failures or malfunctions that would have precluded normal operation. The reason for the lack of left braking action could not be determined.</p>
<p>Probable cause sectie: A loss of left braking action for reasons that could not be determined because postaccident examination did not reveal any evidence of preimpact mechanical failures or malfunctions that would have precluded normal operation, which resulted in a loss of directional control.</p>
<p>Gegenereerde samenvatting: The pilot's inability to maintain directional control during the landing roll due to a lack of left brake action for reasons that could not be determined.</p>

Tabel 11: Voorbeeld van de input tekst, probable cause sectie en gegenereerde tekst die horen bij een ongevalsrapport.

<p>Input tekst: On June 11, 2015, about 0840 mountain standard time, a Beech E33A, N8401N, exited the runway landing surface during the landing rollout at La Cholla Airpark, Tucson, Arizona, and collided with a drainage ditch. The owner/pilot was operating the airplane under the provisions of 14 Code of Federal Regulations Part 91. The commercial pilot was not injured; the airplane sustained substantial damage. The local personal flight departed Tucson, about 0800. Visual meteorological conditions prevailed, and no flight plan had been filed. In a written statement, the pilot reported that the landing touchdown was smooth, with power nearly off. During the landing roll, he applied gentle pressure to the brakes to slow down, however the left side braking action was non-existent, while the right side brake held. The pilot was unable to maintain directional control as the airplane departed the narrow runway and subsequently collided with a drainage ditch and low cement wall. The wreckage was recovered to a secured facility where a Federal Aviation Administration Inspector from the Scottsdale, Arizona, Flight Standards District Office examined the brake system. The inspector reported no evidence of a mechanical malfunction or failure. The hydraulic system was almost empty due to the bleeder valve on the brake caliper was sheared off. The airplane touched down smoothly with the power nearly off. During the landing roll, the commercial pilot applied gentle pressure to the brakes to slow down; however, there was no left braking action, but there was right braking action. The pilot was unable to maintain directional control of the airplane, and it departed the narrow runway and subsequently collided with a drainage ditch and cement wall. During postaccident examination of the brake system, the hydraulic system was found almost empty because the bleeder valve on the brake caliper had sheared off due to impact damage. The examination revealed no evidence of any preimpact mechanical failures or malfunctions that would have precluded normal operation. The reason for the lack of left braking action could not be determined.</p>
<p>Probable cause sectie: A loss of left braking action for reasons that could not be determined because postaccident examination did not reveal any evidence of preimpact mechanical failures or malfunctions that would have precluded normal operation, which resulted in a loss of directional control.</p>
<p>Gegenereerde tekst: The pilot's inability to maintain directional control during the landing roll due to a lack of left brake action for reasons that could not be determined.</p>

Automatische evaluatiescores worden gebruikt om verschillende getrainde modellen met elkaar te vergelijken. Voor de automatische evaluatie van de gegenereerde teksten is gebruik gemaakt van de ROUGE (Recall-Oriented Understudy for Gisting Evaluation) score [31]. De ROUGE score is een set van scores die bedacht is om samenvattingen met elkaar te vergelijken, en een maat te geven van de kwaliteit van de samenvatting ten opzichte van de originele tekst. De scores die hier uitgerekend zijn, zijn de R1, R2 en RL score. De R1 en R2 kunnen gedefinieerd worden als de Rn scores, waar de n staat voor het aantal overlappende n-grams tussen de teksten. Een n-gram kan worden gedefinieerd als een set van n aangrenzende woorden. Bijvoorbeeld, voor ‘excessive breaking caused the accident’ zijn de 1- of unigrams ‘excessive’, ‘breaking’, ‘caused’, ‘the’ en ‘accident’. De 2- of bigrams zijn ‘excessive breaking’, ‘breaking caused’, ‘caused the’ en ‘the accident’. De Rn score kan worden uitgedrukt als:

$$Rn_{precision} = \frac{\sum \text{overlappende } n - \text{grams}}{\sum n - \text{grams gegenereerde tekst}} \cdot 100$$

$$Rn_{recall} = \frac{\sum \text{overlappende } n - \text{grams}}{\sum n - \text{grams referentie tekst}} \cdot 100$$

Bij de RL score wordt gebruik gemaakt van de langste overeenkomende reeks van woorden tussen de referentie tekst en de gegenereerde tekst. De ROUGE score kan worden verduidelijkt aan de hand van een voorbeeld. Een mogelijke referentie en gegenereerde tekst zouden kunnen zijn: “The pilot's failure to maintain clearance from a powerline during an intentional low altitude maneuver.” (referentie), en “The pilot's failure to maintain clearance from a powerline.” (gegenereerd). De R1 recall score kan berekend worden voor dit enkele voorbeeld. In dit geval bevat de referentie tekst vijftien woorden en de gegenereerde tekst negen woorden. Tussen de twee teksten zijn er negen woorden die overlappen. De R1 recall score is dus gelijk aan $R1 = 9/15 \cdot 100 = 60$ en de R1 precision score is gelijk aan $R1 = 9/9 \cdot 100 = 100$

De ROUGE scores voor het BART-large model toegepast op de test set zijn te zien in Tabel 12. Om de ROUGE scores over een gehele dataset te berekenen wordt de score gemiddeld over alle voorbeelden in de dataset. Voor de R1 recall score is in Tabel 12 een waarde van 0.494 gegeven. Dit betekent dat gemiddeld de helft van de woorden die voorkomt in een gegenereerde tekst, ook voorkomt in de bijhorende referentietekst.

Tabel 12: ROUGE scores van het model toegepast op de testset.

	R1	R2	RL
Precision	59.7	39.0	52.8
Recall	49.4	32.1	43.8

De ROUGE scores die hier behaald worden zijn relatief hoog in vergelijking met scores die gehaald worden bij andere samenvattingstaken. Zo behaalt het BART-model R1, R2 en RL recall scores van respectievelijk 44.16, 21.28 en 40.90 op de CNN/DailyMail dataset, een dataset die vaak wordt gebruikt om samenvattingsalgoritmes met elkaar te vergelijken (Lewis, et al., 2019 [30]). Een reden voor het feit dat de ROUGE scores relatief hoog zijn voor het probable cause model kan zijn dat de probable cause teksten relatief homogeen zijn in woordgebruik en inhoud. In Tabel 8 is bijvoorbeeld te zien dat woorden als ‘pilot’, ‘failure’, ‘directional’ en ‘control’ vaak voorkomen. Omdat ROUGE scores niet een directe maat geven voor inhoudelijke correctheid is het niet mogelijk om aan de hand van deze scores een sluitend oordeel te geven over de uiteindelijke toepasbaarheid van het model. Wel kan de ROUGE score worden gebruikt als objectieve score om verschillende modellen met elkaar te vergelijken in een volgende fase van het onderzoek.

4.5.7 Conclusie

In dit onderzoek is een BART-large model getraind en toegepast voor het genereren van een samenvatting van de 'probable cause' van een voorvalrapport in een of twee zinnen op basis van een tekst van maximaal circa 800 woorden waarin de oorzaak van een ongeval beschreven staat. Het model is in staat zinnen te genereren die taalkundig correct en begrijpelijk zijn. Om een goed oordeel te kunnen geven over de juistheid en kwaliteit van de gegenereerde samenvatting van de probable cause, en in welke mate deze samenvatting geschikt is om op grote schaal analyses te doen, moet verder onderzoek worden gedaan.

4.5.8 Aanbevelingen

De volgende aanbevelingen worden gedaan:

- De ongevalsrapporten van de NTSB volgen een vaste structuur waarbij de oorzaak van het ongeval wordt omschreven in een aparte sectie. Dit is niet altijd het geval bij ongevals- en incidentenrapporten van andere onderzoeksinstanties, zoals bijvoorbeeld in rapporten van de AAIB of in CADORS. Daarom wordt aangeraden om het model toe te passen op rapporten uit andere databronnen dan de NTSB om te onderzoeken hoe het model in dat geval presteert. Dit geeft informatie over de inzetbaarheid van het model op verschillende soorten rapporten (verschillend in opbouw, stijl, formulering etc.)
- Om meer inzicht te halen uit de gegenereerde teksten zou een topic model toegepast kunnen worden. Topic modelling is een unsupervised Machine Learning techniek waarbij de onderwerpen onderliggend aan een set van teksten achterhaald kunnen worden. Dit kan vervolgens worden gebruikt om onderwerpen toe te kennen aan individuele teksten, om zodoende statistieken te genereren van een gehele corpus van input teksten. Op deze manier kan bijvoorbeeld onderzocht worden hoe het optreden van bepaalde oorzaken verandert in de tijd of wat het verschil is in onderliggende oorzaken in voorvallen in de commerciële luchtvaart ten opzichte van de kleine luchtvaart.
- Recentelijk zijn er meer modellen ontwikkeld die vergelijkbaar zijn met het BART-model. Een nieuw model is het T5 model dat uitgebracht is door onderzoekers van Google [33]. Dit model heeft veel aandacht gekregen in de onderzoeksgemeenschap. De aanbeveling is om de toepassing en prestatie van dit model op deze case te onderzoeken en te vergelijken met het BART-model.
- Scores zoals ROUGE die automatisch worden berekend door een computer hebben een aantal nadelen. Een voorbeeld van zo'n nadeel is dat het mogelijk is om een twee samenvattingen te schrijven in andere bewoordingen, met een verschillend aantal woorden, maar die wel beide inhoudelijk correct zijn. In dit geval kan de ROUGE score toch verschillend zijn voor de twee teksten, wat niet wenselijk is omdat het met het oog op de uiteindelijke toepassing vooral belangrijk is dat de inhoud van de samenvatting correct is. Daarom is het noodzakelijk om het model te laten evalueren door een veiligheidsanalist. Een mogelijke score die daarbij gehanteerd kan worden is de 'correctness' of correctheid zoals gedefinieerd in [27]. Hierbij krijgt een gegenereerde tekst de score 1 als de inhoud fout of misleidend is, 2 als de inhoud gedeeltelijk correct is en 3 als de inhoud voornamelijk correct is. Een indicatie van de prestaties van het model zou dan gegeven kunnen worden door het gemiddelde te nemen van de correctheid van alle geëvalueerde teksten. Naast de correctheid kan onderzoek gedaan worden naar welke stukken informatie gegeven worden in een gegenereerde tekst. Denk hierbij bijvoorbeeld aan de oorzaak van een ongeval, de relevante actoren of in welke vluchtfase het ongeval heeft plaatsgevonden. Deze beoordeling op correctheid is een noodzakelijke aanvulling op de ROUGE scores om de uiteindelijke prestatie van het model te vast te stellen.

4.6 Case 3: Sentiment analyse op sociale media berichten

4.6.1 Inleiding

Deze paragraaf beschrijft het toepassen van sentiment analyse op openbare tekstdata. Onder openbare bronnen worden bronnen verstaan die publiekelijk toegankelijk zijn en zich niet bevinden in een beveiligde omgeving waar expliciete toegang voor moet worden verleend. In eerste instantie zou deze derde use case zich richten op het uitvoeren van een sentiment analyse op tekstdata van openbare forums over onderwerpen zoals luchtvaartveiligheid en aan luchtvaart gerelateerde risico's. Echter, een inventarisatie van de beschikbare data op openbare forums met betrekking tot deze onderwerpen leverde (te) weinig data voor een sentiment analyse. Tekstdata met publieke opinies over luchtvaartveiligheid en risico's bleek in mindere mate aanwezig dan in eerste instantie werd verwacht. Er is voornamelijk gezocht op forums waar veel general aviation en drone vliegers te vinden zijn. Daar zijn weinig posts (berichten) gevonden die gingen over luchtvaartveiligheid en risico's. De posts die wel zijn aangetroffen over deze onderwerpen werden vaak geplaatst door een klein groepje gebruikers en werden ook niet vaak bekeken. Daarnaast zijn er weinig uitgesproken meningen in de berichten aangetroffen op deze forums (de berichten hebben voornamelijk een neutraal sentiment). Tenslotte zou het verzamelen van data van deze fora leiden tot een grote werklast aangezien voor elk forum een apart op maat gemaakt methode zou moeten worden ontwikkeld om de data te verzamelen².

Andere bronnen waarin grote hoeveelheden openbare tekstdata kunnen worden gevonden zijn social media platforms zoals Facebook en Twitter. Met de opkomst van sociale media platforms is de hoeveelheid beschikbare openbare tekstdata drastisch toegenomen. Vaak bieden deze platforms ook methodes om (tekst)data grootschalig te verzamelen voor analyse (soms tegen betaling), waardoor het eenvoudig is om data regulier te verzamelen. Bovendien is de kwaliteit van de data goed aangezien het direct van de bron wordt verkregen. Om deze redenen is ervoor gekozen om tekstdata van Twitter te verzamelen voor deze use case.

4.6.2 Definitie van sentiment analyse

Er zijn verschillende definities van sentiment analyse. In deze use case is gekozen voor de volgende definitie: "Sentiment analysis: An examination that aims to determine the attitude of a speaker or a writer on a given topic, which may be his or her "judgment" or "evaluation" of that topic and/or "emotional state" connected to that topic." (Nisbet, Elder & Miner, 2009 [35]). Een belangrijke reden om te kiezen voor deze definitie is de nuance die wordt gelegd op het sentiment dat ofwel gaat over het onderwerp en/of over de emotionele staat van de schrijver gerelateerd aan het onderwerp ("which may be his or her "judgment" or "evaluation" of that topic and/or "emotional state" connected to that topic."). Dit deel van de definitie geeft namelijk aan dat het sentiment van een stuk tekst niet een gegeven is voor het sentiment dat heerst over dat onderwerp, maar dat het sentiment ook gerelateerd kan zijn aan het onderwerp. Een voorbeeld: Stel dat uit sentiment analyse blijkt dat 90% van de teksten met als onderwerp coronavaccinatie als negatief zijn geclassificeerd. Dat hoeft echter niet te betekenen dat de schrijvers van de teksten tegen coronavaccinatie zijn. Het kan ook zijn dat de schrijvers vóór coronavaccinatie zijn, maar ontevreden zijn over de lange wachttijden voor de vaccinatie. Kortom, nadat een sentiment analyse is uitgevoerd moet de achterliggende reden voor het negatieve of positieve sentiment verder worden uitgezocht. In feite is de sentiment analyse die in deze

² Bovendien staat niet elk forum toe dat data op een programmatische manier wordt vergaard (i.e. web scraping).

case wordt uitgevoerd een sentiment classificatie van teksten, en niet zozeer een analyse van het sentiment en onderliggende oorzaken.

4.6.3 Data

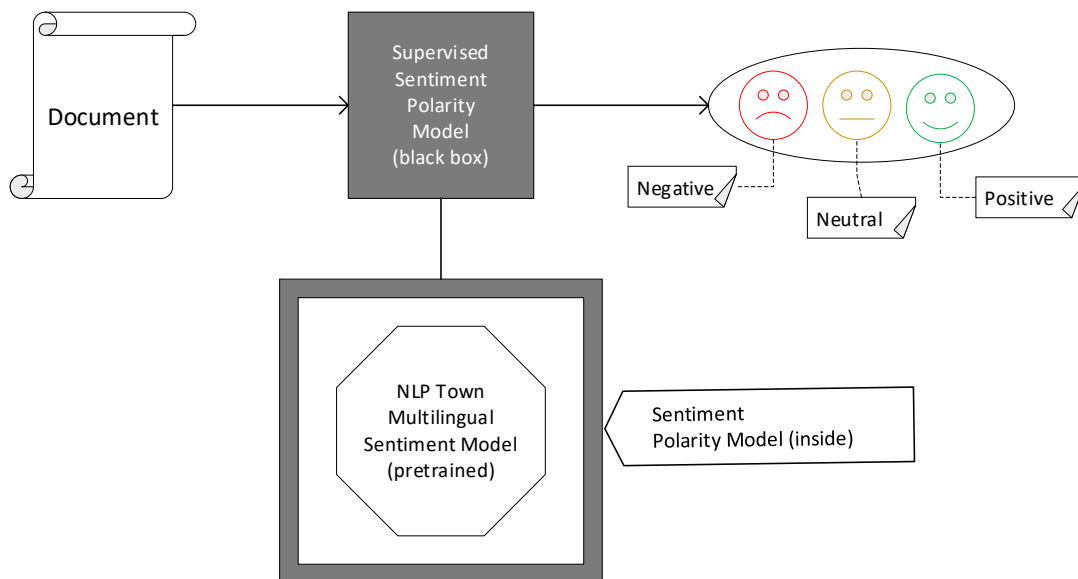
De tekstdata die wordt gebruikt in deze use case is afkomstig van het social media platform Twitter. Twitter staat bekend om de korte berichten (maximaal 280 karakters) die gebruikers kunnen plaatsen, ook wel Tweets genoemd. Deze Tweets worden vaak gebruikt om kort en bondig informatie te verstrekken en is populair onder een brede groep Nederlanders. Dit maakt Twitter een geschikte kandidaat voor het verzamelen van sentimenten uit alle lagen van de samenleving. Het verzamelen van Tweets kan op een eenvoudige manier worden uitgevoerd via de officiële Twitter API. Om gebruik te kunnen maken van de Twitter API moet de gebruiker een API sleutel hebben. Deze sleutel dient als authenticatie en verleent toegang tot Twitter data zoals Tweets. Om een API sleutel te ontvangen moet de gebruiker een Twitter Developer Account aanvragen bij Twitter. Met deze gratis account en API sleutel kunnen Tweets tot 7 dagen geleden worden verzameld. Na het verkrijgen van de API sleutel zijn er op wekelijkse basis Tweets verzameld met een geautomatiseerd script. De Tweets die zijn verzameld zijn geselecteerd op basis van een keyword search op verschillende luchtvaarttermen in de Twitter dataset. De complete lijst met keywords waarop is gezocht is opgenomen in Appendix D. Tussen 26 oktober 2020 en 06 december 2020 zijn er in totaal 49.050 Tweets verzameld voor deze studie (gemiddeld 8.175 Tweets per week).

4.6.4 Model 1: Supervised sentiment analyse

Supervised sentiment analyse is een supervised tekstclassificatiemodel dat wordt getraind met historische tekstdata, waaraan reeds een sentimentlabel is toegekend. Er zijn verschillende manieren om aan stukken tekst een sentimentlabel toe te kennen:

- Een analist beoordeelt de stukken tekst zelf en labelt handmatig de tekst met een sentiment- of polariteitscore.
- De schrijver van de tekst kent zelf een sentiment- of polariteitscore toe aan de tekst. Voor sommige stukken tekst wordt de schrijver gevraagd om zelf een score toe te kennen, bijvoorbeeld bij product of diensten reviews. Deze scores kunnen gebruikt worden om de polariteit in te schatten van de bijbehorende tekst. Zo kunnen we aannemen dat het sentiment van een tekst over een product review waarbij de schrijver een 4 (op een schaal van 0-10) voor het product heeft gegeven negatief is.

Bij het gebruik van een supervised sentiment analysemodel is het van belang dat er voldoende trainingsdata beschikbaar is om het model te trainen. Omdat er voor deze use case onvoldoende tijd was om zelf trainingsdata te genereren, is er gebruik gemaakt van een model dat al was getraind om de polariteit te voorspellen van teksten in de Nederlandse taal. Het model dat gebruikt is voor de supervised sentiment analyse is een BERT-model van NLP Town [29]. Dit model is getraind op basis van ongeveer 80.000 Nederlandse product reviews van Amazon.com. De hoeveelheid sterren (1-5 sterren) per productreview is gebruikt als maat voor de polariteit van de product review.



Figuur 16: Proces voor supervised sentiment analyse.

4.6.5 Model 2: Unsupervised sentiment analyse

In deze use case is gebruik gemaakt van een lexicon-based unsupervised sentiment analysemodel. De lexicon-based sentiment analyse maakt gebruik van lexicons, ook wel woordenlijsten genoemd. Deze lexicons bevatten verschillende woorden waarvan de polariteit per woord is beoordeeld door één of meerdere personen. Het beoordelen van de polariteit kan op verschillende manieren worden uitgevoerd, waarbij het gebruikelijk is om een grove scoring te gebruiken, zoals bijvoorbeeld in Tabel 13.

Tabel 13: Sentiment scoring methode.

Score	Score beschrijving
++	Zeer positief
+	Positief
0	Neutraal
-	Negatief
--	Zeer negatief

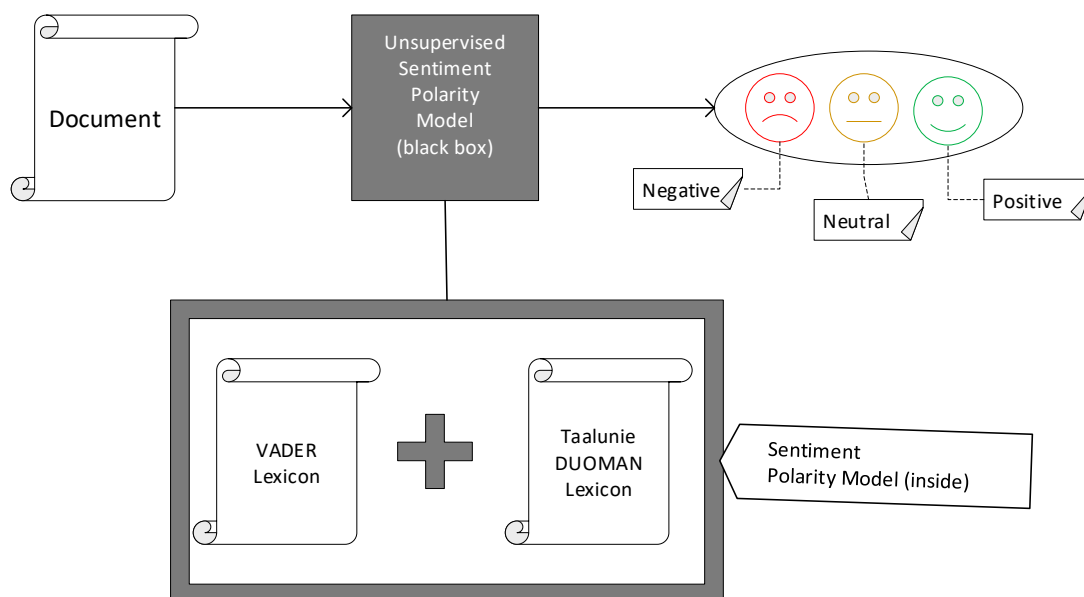
De lexicons die gebruikt worden voor sentiment analyse dienen een grote hoeveelheid woorden te bevatten. Woorden die wel in de te analyseren tekst staan, maar niet in het lexicon, kunnen immers niet gescoord worden. Deze use case heeft gebruik gemaakt van een combinatie van twee lexicons, namelijk:

1. VADER [36]. Het VADER-model maakt gebruik van een uitgebreid Engels lexicon (11.000 woorden) dat gespecialiseerd is in het analyseren van teksten afkomstig van social media. Het lexicon bevat daarom naast een grote hoeveelheid Engelstalige termen, ook emoticons, straattaal en afkortingen. Ondanks dat het VADER lexicon voornamelijk Engelse termen bevat is het model geschikt voor deze use case aangezien in Nederlandse social media berichten soms ook Engelse taal wordt gebruikt, evenals emoticons, Engelstalige straattaal en Engelstalige afkortingen.
2. Duoman Subjectivity lexicon [38]. Om zo veel mogelijk Nederlandstalige teksten te kunnen beoordelen op sentiment is er ook gebruik gemaakt van een Nederlandstalig lexicon. Dit is het Duoman Subjectivity lexicon, gepubliceerd door de Nederlandse Taalunie. Dit lexicon omvat 9.000 Nederlandse woorden die zijn

beoordeeld op sentiment. Ter illustratie is een deel van dit lexicon in de onderstaande tabel opgenomen. In de “Score” kolommen staat het sentiment van het woord zoals dat beoordeeld is volgens de scoringsmethodiek uit Tabel 13. Voor sommige woorden uit het lexicon zijn er twee of meer scores gegeven. Deze woorden zijn door meerdere analisten gescoord. Mochten er twee of meer scores beschikbaar zijn dan wordt het gemiddelde van de scores per woord gebruikt volgens de numerieke score zoals weergegeven in Tabel 13.

Woord	Score 1	Score 2	Score ...
autodidact	+	n.v.t	
autorijden	0	n.v.t	
autoritair	+	--	
autoriteit	+ +	+	
autosport	0	n.v.t	
avanceren	+	n.v.t	
averechts	--	n.v.t	
aversie	--	n.v.t	

In onderstaand figuur wordt het proces voor het bepalen van sentiment met een unsupervised lexicon methode weergegeven. Het document wordt in dit project gezien als een Tweet. In het sentiment polarity model wordt vervolgens de Tweet woord voor woord beoordeeld op sentiment met behulp van zowel het VADER-lexicon als het Duoman lexicon. Alle scores van het stuk tekst worden samengevoegd voor een totale sentiment score tussen de -1 en de +1. Hierbij is -1 het meest negatief en +1 het meest positief.



Figuur 17: Proces voor unsupervised sentiment analyse.

4.6.6 Resultaten

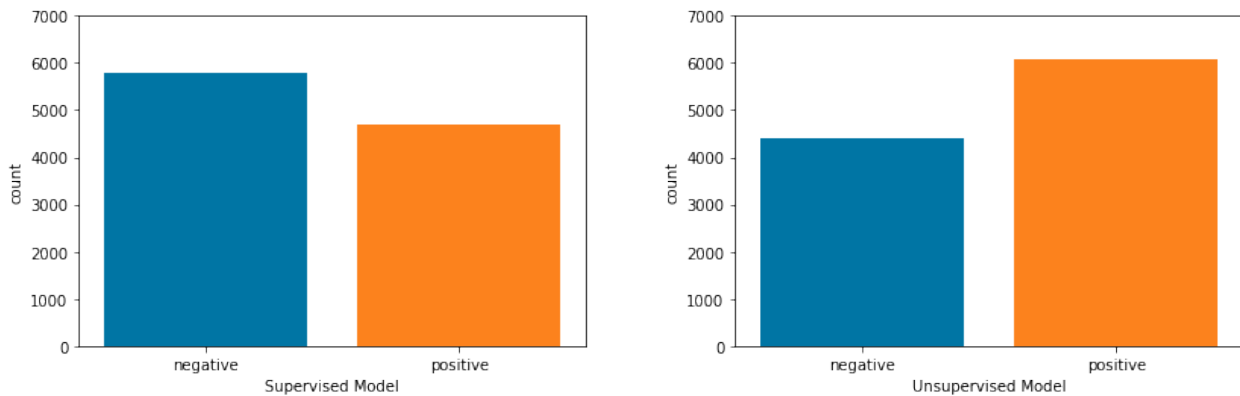
Een veelgebruikte methode om snel inzicht te krijgen in een dataset met veel tekst is het genereren van een wordcloud. Een wordcloud is een figuur waarin de meest gebruikte woorden uit een dataset zijn opgenomen. Hoe vaker een woord is gebruikt in de dataset, hoe groter het woord (letterformaat) in de wordcloud wordt afgebeeld. In Figuur 18 is een wordcloud opgenomen van de Twitter dataset die is verkregen voor deze studie (zie vorige paragraaf). Deze wordcloud bevat de top-1000 van meest gebruikte woorden, exclusief typisch Nederlandse stopwoorden die over het algemeen het vaakst voorkomen maar verder geen betekenis hebben. Wat opvalt in deze wordcloud zijn onder andere:

- De grote populariteit van het onderwerp Lelystad Airport;
- Operationele partijen zoals Schiphol en KLM;
- Gebruik van signaalwoorden die kunnen duiden op een vraagstelling, onbegrip of vraag om uitleg, zoals waarom, mensen, gewoon.
- Corona.



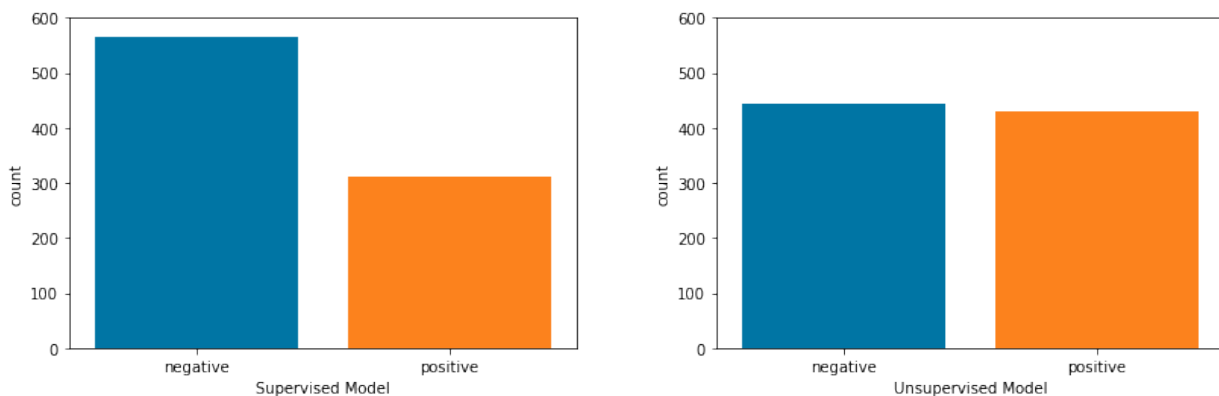
Figuur 18: Wordcloud van de verzameling Tweets in case 3.

Na de vergaring van de Tweets zijn de unsupervised en supervised methodes toegepast op de Tweets dataset. In totaal zijn 11.540 unieke Tweets gelabeld met een positief of negatief sentiment label. De overige Tweets in de dataset zijn niet gelabeld omdat ze ofwel geen unieke tekst bevatten maar alleen een re-Tweet waren of omdat het een Tweet was die dubbel is verzameld. In de onderstaande figuren staat de totale hoeveelheid Tweets per polariteit per sentiment analyse methode (model 1 – supervised (BERT), model 2 - unsupervised). Wat opvalt is het verschil in ratio tussen positief en negatief aangemerkte Tweets tussen het supervised en unsupervised model. Het supervised model markeert het merendeel van de Tweets als negatief, terwijl het unsupervised model in dezelfde set meer Tweets een positief label geeft. De reden voor deze verschillen in labelling is nog onbekend.

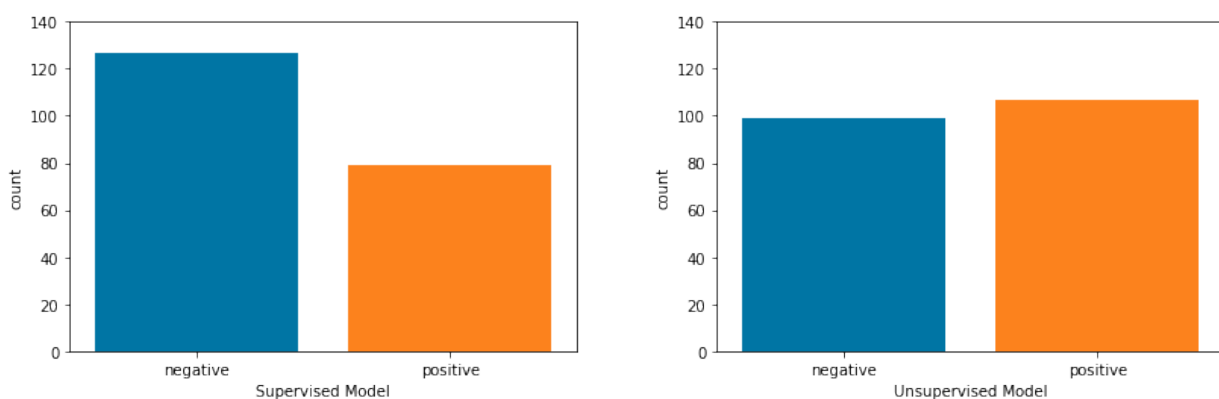


Figuur 19: Resultaat van sentiment analyse voor alle Tweets, vergelijking tussen supervised en unsupervised model.

De figuren hieronder laten eveneens de polariteit zien, maar dan voor Tweets waarvan in de tekst bepaalde termen voorkomen, in dit geval “Lelystad” (Figuur 20) en “economie” (Figuur 21).



Figuur 20: Resultaat van sentiment analyse, vergelijking tussen supervised en unsupervised model, specifiek voor het onderwerp “Lelystad”.



Figuur 21: Resultaat van sentiment analyse, vergelijking tussen supervised en unsupervised model, specifiek voor het onderwerp “economie”.

De volgende twee afbeeldingen zijn voorbeelden van de Tweets waarvan het sentiment is geclassificeerd. In de eerste Tweet spreekt de schrijver zijn steun uit voor de luchtvaart, ondanks dat hij het geluid van vliegtuigen hoort. Beide modellen hebben deze Tweet als positief aangemerkt.

Genieten, ik woon nu in Adam centrum en mis het geluid van vliegtuigen.
 Vergeet niet dat Schiphol lokaal veel geld oplevert, werkgelegenheid creëert. En mocht u ooit gaan verkopen ,dan vangt u mede door Schiphol de hoofdprijs.
 Misschien wel de meest gewilde plek in Nederland

[Translate Tweet](#)

2:40 PM · Nov 14, 2020 · Twitter Web App

1 Like

De volgende Tweet is door beide modellen als negatief gelabeld. De schrijver van de Tweet is kritisch op het geluid dat wordt geproduceerd door de luchtvaart.

Ik ook. Ik ben hier geboren. En het geluid van toen is n  t te vergelijken met dat van nu. We worden kapot gemaakt en genegeerd door Schiphol

[Translate Tweet](#)

1:47 PM · Nov 14, 2020 · Twitter Web App

Onderstaand voorbeeld is een Tweet die door beide modellen verschillend is gelabeld. Het supervised model heeft deze Tweet als negatief gelabeld en het unsupervised model als positief.

Einde KLM; einde Schiphok; einde luchthaven Lelystad; 1.700.000.000 euro weggegooid aandelen en reeds verstrekte subsidie; einde toeleveranciers; einde werkgelegenheid. Kortom miljarden kapitaalvernietiging. Hoezo economisch beleid.

6:36 PM · Oct 31, 2020 · Twitter for iPhone

Uit de bovenstaande figuren is op te maken dat de modellen niet altijd dezelfde uitkomst voorspellen voor dezelfde Tweets. Daarnaast is het momenteel niet mogelijk om de complete dataset automatisch te evalueren op correcte labelling omdat de Tweets van zichzelf geen sentimentindicatie hebben (ze zijn niet gelabeld). Wel zijn er met behulp van een korte inventarisatie van de resultaten een aantal bevindingen opgesteld:

- Stukken tekst die satirische uitspraken bevatten zijn niet of moeilijk als zodanig te herkennen door beide modellen.
- Het supervised model classificeert Tweets over het algemeen vaker als negatief dan het unsupervised model.
- Korte Tweets zijn moeilijker te classificeren omdat ze minder lingu  stische eigenschappen bevatten dan langere Tweets.

- Er zijn Tweets waarbij geen enkel woord uit de tekst in een van de twee gebruikte lexicons van het unsupervised model zijn terug te vinden. Dit resulteerde vaak in een Tweet met sentiment score 0, wat werd gelabeld als positief. Dit is een beperking van de huidige lexicons en daarmee het unsupervised model.

Beide modellen zijn ook toegepast op andere datasets om hun prestaties te kunnen evalueren. Deze datasets waren op voorhand al gelabeld met positieve en negatieve sentiment labels waardoor de twee gebruikte modellen eenvoudig gecontroleerd kunnen worden op correctheid van toegekende sentimenten. De eerste dataset waarop de modellen zijn toegepast is de dataset die in een van de cases uit de literatuurstudie is beschreven (zie Appendix C.5.5). Deze case probeerde het sentiment van Engelstalige Tweets over verschillende Amerikaanse luchtvaartmaatschappijen te voorspellen. Deze dataset is geschikt om de twee modellen te evalueren omdat deze dataset eveneens gebruikt is in een sentiment analyse gebaseerd op Tweets gerelateerd aan de luchtvaart. Na het toepassen van de modellen is het percentage correct voorspelde sentimenten berekend over de totale hoeveelheid toegekende sentimenten. Het supervised model behaalde hier een percentage van 77,5% juiste voorspellingen. Het unsupervised model behaalde een percentage van 55,5% correcte voorspellingen van het sentiment.

Een tweede externe dataset waarop beide modellen zijn toegepast is een Nederlandstalige boekreview dataset. Deze boekreview dataset is minder representatief voor de use case 3 omdat het geen tekstdata bevat van een social media platform zoals Twitter. Aan de andere kant is deze dataset wel bruikbaar voor een evaluatie van de modellen omdat het een Nederlandstalige dataset is met 14.000 Tweets die al gelabeld zijn, wat een voldoende grote dataset biedt om een evaluatie uit te kunnen voeren. Na het toepassen van de modellen op deze dataset is het percentage correct voorspelde sentimenten berekend over de totale hoeveelheid voorspelde sentimenten. Het supervised model behaalde hier een percentage van 70% correcte voorspellingen. Het unsupervised model behaalde een percentage van 60% correcte voorspellingen.

4.6.7 Conclusies

Met behulp van sentiment analyse is het mogelijk om het sentiment van verschillende stukken tekst te bepalen. Een gedegen sentiment analyse houdt meer in dan enkel het voorspellen (labelen) van het sentiment. Een daaropvolgende analyse van de resultaten is essentieel om te bepalen hoe accuraat de voorspelde sentimenten zijn. Daarnaast is het belangrijk om met behulp van andere analysetechnieken de verkregen sentimenten verder te onderzoeken om zo beter te kunnen vaststellen over welke onderwerpen welke sentimenten heersen.

De techniek voor sentiment analyse is snel en eenvoudig toe te passen. Er is zowel een supervised als een unsupervised methode beschikbaar die gebruikt kan worden met relatief weinig ervaring. Echter, het is op dit moment niet duidelijk hoe accuraat beide modellen presteren in de huidige case. Daarvoor is een uitgebreide analyse van de resultaten noodzakelijk. Het proces dat is opgesteld om regulier Tweets te verzamelen kan eenvoudig worden opgesteld en gecontinueerd. Dat levert in de nabije toekomst een grotere dataset op voor sentiment analyse.

Uit de toepassing van sentiment analyse op Tweets over luchtvaart gerelateerde onderwerpen wordt duidelijk dat een dergelijke analyse geschikt is om de polariteit vast te stellen over verschillende onderwerpen aan de hand van Tweets. Tweets lijken vaker negatieve dan positieve sentimenten te bevatten wat in ogenschouw genomen moet worden bij de analyse van resultaten. We merken op dat de databron waarvoor het sentiment wordt voorspeld van invloed is op (de interpretatie van) de resultaten. Het kan zijn dat bepaalde databronnen een verzamelplaats zijn voor mensen met

een bepaald standpunt of sentiment. Dit kan de hoeveelheid positieve of negatieve sentimenten over een bepaald onderwerp in de databron beïnvloeden.

In deze use case is gebleken dat beide modellen in sommige gevallen verschillende resultaten geven voor dezelfde Tweets. Het automatisch evalueren van de modellen op correcte toekenning van sentimenten was niet mogelijk in deze studie omdat de dataset met Tweets niet het daadwerkelijke sentiment per Tweet bevatte. Daarom is het alleen mogelijk om handmatig vast te stellen of een model een correcte voorspelling doet. Met behulp van andere datasets die online beschikbaar waren voor sentiment analyse is een eerste poging gedaan om de modellen te toetsen op correctheid. Deze andere datasets bevatten het daadwerkelijke sentiment per stuk tekst. Op basis van deze datasets bleek dat het supervised model beter presteerde dan het unsupervised model.

Het supervised model leent zich het beste voor verbetering aangezien de voorspellingen die het model doet afhankelijk zijn van de historische data waarop het model is getraind. Het is daarom mogelijk om een sentiment analysemodel te trainen dat gespecialiseerd is in sentiment analyse van luchtvaart gerelateerde onderwerpen. Daarnaast kunnen aspecten van het model geïdentificeerd worden waar nog niet voldoende op wordt gepresteerd. Daarmee kunnen de ontwikkelaars het model met specifieke trainingsdata verder finetunen. Dit zal de resultaten ten goede komen. Een voorwaarde voor het her-trainen is dat gelabelde historische data beschikbaar zijn. Een voordeel van deze methode is dat het model niet alleen getraind, maar ook getest kan worden met deze historische data. Het nadeel is dat het verzamelen en genereren van historische data voor dit soort taken vaak arbeidsintensief is.

In het kader van de NVA kan het Ministerie van IenW een sentiment analyse inzetten voor de identificatie en analyse van het sentiment over een bepaald onderwerp met betrekking tot luchtvaartveiligheid of risico's. De resultaten van een sentiment analyse kunnen vervolgens benut worden voor veiligheidspromotie en communicatie over de NVA.

4.6.8 Aanbevelingen

De volgende aanbevelingen worden gedaan:

- Het verder onderzoeken en evalueren van de supervised sentiment analysemodel (BERT-model). Daarbij kan het model ook worden toegepast op andere datasets dan Tweets, bijvoorbeeld op ingevulde vragenlijsten (questionnaires), of teksten met verschillende kwaliteit, of met meer evenwichtige sentimenten (minder 'zwart/wit' sentiment zoals vaak in Tweets wordt aangetroffen).
- Andere preprocessing methodes onderzoeken die het trainen van het model kunnen verbeteren. Het verbeteren van spelling en het toevoegen van Nederlandse straattaal en afkortingen in lexicons zijn twee zaken die kunnen worden uitgevoerd. Tevens kan het lexicon dat gebruikt wordt in de unsupervised methode worden uitgebreid.
- Het identificeren en toepassen van andere analysemethodes waarmee de resultaten van een sentiment analyse kunnen worden geanalyseerd. Dat moet tot meer inzicht leiden in de sentimenten, het precieze onderwerp van het sentiment, en de redenen voor het sentiment.
- Ontwikkelen van interactieve visualisatietechnieken om periodieke resultaten van de sentiment analyse te kunnen presenteren en analyseren.

5 Conclusies en aanbevelingen

5.1 Conclusies

In het kader van het Kennis voor Beleid programma heeft het NLR in 2020 een onderzoek uitgevoerd naar het gebruik van veiligheids- en operationele data ten behoeve van de nationale veiligheidsanalyse (NVA). Het eerste doel was het identificeren en het beoordelen van databronnen die gebruikt kunnen worden voor een nationale veiligheidsanalyse, in het bijzonder voor gevaarsidentificatie, risicobeoordeling en risicomitigatie. Het tweede doel van dit onderzoek was het inventariseren van welke verrijkingen en normalisaties van data mogelijk zijn. Het derde doel was het verkennen van de mogelijkheden van Machine Learning technieken om de NVA te ondersteunen. Op basis van het uitgevoerde onderzoek worden de volgende conclusies getrokken over deze drie onderzoeksdoelen.

Databronnen

- De lijst met databronnen en contextuele data kan als basis dienen voor de NVA. Er is een aanzienlijke hoeveelheid databronnen beschikbaar voor de NVA, zowel voor gevaarsidentificatie, risicobeoordeling en risicomitigatie. De kwaliteit en betrouwbaarheid van de data en informatie in de bronnen varieert sterk. Er zijn in deze studie een aantal criteria geformuleerd, als leidraad voor de analist, om de kwaliteit van de data(bron) voor het betreffende criterium te beoordelen.
- De dekking van databronnen voor Commercial Air Transport (CAT) is een stuk hoger dan voor de domeinen General Aviation (GA) en drone operations: 98% van de databronnen beslaat het CAT-domein, tegenover 65% voor GA, en 45% voor drones.
- De databronnen dekken alle elementen in het luchtvaartstelsysteemmodel af. Het merendeel van de bronnen geeft informatie over de (inrichting van de) vliegoperatie en de prestatie van de operatie. Ongeveer één derde van de databronnen biedt informatie over de systeemelementen die betrekking hebben op het toezicht, en het leren en bijsturen binnen het systeem.
- De sectorpartijen beschikken over databronnen die ruwe (niet verwerkte of geanalyseerde) data bevatten, zoals vluchtdata, radardata, en voorvalmeldingen. Deze data zijn zeer nuttig ter ondersteuning van een NVA. Echter, het vraagt een aanzienlijke inspanning en middelen om deze data te verwerken, op te slaan, en te analyseren voor een risicoanalyse. Daarom wordt aanbevolen om in de NVA-methode richtlijnen op te stellen voor welk soort vraagstukken het de moeite waard is om deze ruwe data te benutten.
- De overheid (zoals Ministerie IenW, ILT, ABL) beschikt met name over databronnen die statistiek en analyses bevatten. Deze informatie is eenvoudiger toegankelijk en laagdrempelig te gebruiken in een NVA in vergelijking met ruwe data. Echter, het gebruik van bronnen met statistiek en analyses in de NVA moet zorgvuldig gebeuren omdat in de onderliggende bronnen aannames, keuzes en interpretaties zijn gemaakt die invloed hebben op de toepasbaarheid of representativiteit van de bronnen voor de NVA. Bovendien zijn data uit verschillende statistieken meestal lastig te combineren vanwege de verschillende contexten en dimensies, en onduidelijkheid over overlap van onderliggende datapunten.

Verrijking en normalisatie van data

Er zijn veel soorten data beschikbaar in databronnen die benut kunnen worden in de analyse van gevaren en risico's, en die context kunnen bieden in dergelijke analyses. Deze studie geeft een overzicht van 76 contextuele data die kunnen helpen om data te verrijken en te normaliseren. Deze contextuele data bieden informatie over de mate van activiteit of productie van een partij, en/of informatie over het 'systeem' en de omgeving waarin de

luchtvaartoperatie plaatsvindt. In deze studie is het gebruik van verrijking en normalisatie van data geïllustreerd aan de hand van een aantal voorbeelden.

Toepassing van Machine Learning

Gelet op de reikwijdte van ML-technieken en mogelijke toepassingen is in dit onderzoek gekozen voor de toepassing van text-mining technieken ter ondersteuning van de NVA. Het NLR-team heeft enkele ML-technieken op het gebied van text-mining toegepast op drie use cases om de haalbaarheid en bruikbaarheid van deze technieken te evalueren. Naar aanleiding van deze toepassing hebben we per case een aantal conclusies en aanbevelingen opgesteld. Tevens kunnen we enkele algemene conclusies trekken over de toepassing van ML ten behoeve van veiligheidsanalyses:

- Op basis van de drie cases studies en de resultaten van de literatuurstudie lijken ML-technieken goede mogelijkheden te bieden voor ondersteuning van veiligheidsanalyses. De technieken lenen zich vooral voor het analyseren van grote datasets en het voorspellen van bepaalde uitkomsten op basis van historische data.
- Het toepassen van ML vraagt naast kennis van 'data science' ook domeinkennis van luchtvaart(veiligheid) om zowel de ML-modellen te ontwikkelen, te trainen en te testen, maar ook om de resultaten op waarde te kunnen beoordelen.
- Het finetunen van de modellen en het maximaliseren van de opbrengst van de ML-technieken kost veel tijd. De pre-processing en finetuning spelen een belangrijke rol in het analyseproces en leggen een aanzienlijk beslag op de capaciteit. Het ligt daarom voor de hand om ML vooral in te zetten daar waar het de efficiëntie kan verhogen, bijvoorbeeld voor het vervangen van regulier handmatig classificeren van data. De inzet van ML-technieken voor een eenmalige case weegt waarschijnlijk niet op tegen de inspanning die de analist moet verrichten om een dataset voor te bereiden, een model te ontwikkelen en te trainen, en vervolgens toe te passen en te valideren.
- De ontwikkeling en training van een model wordt beter en gaat sneller als de analist beschikt over een gelabelde dataset. De bruikbaarheid van trainingsdata van andere bronnen of uit andere toepassingen was beperkt in dit onderzoek door afwijkingen in datastructuur, type, inhoud en/of format. De analist kan in deze situatie kiezen om een eigen trainingset ontwikkelen of een handmatige validatie achteraf uit voeren op de resultaten. Het gebrek aan gelabelde data in dit onderzoek betekent dat de validatie van de resultaten op dit moment handwerk is en veel inspanning vraagt.

5.2 Aanbevelingen

We hebben in dit onderzoek een aantal aanbevelingen opgesteld voor de verdere ontwikkeling van de toepassing van ML-technieken in de drie cases. De gebruikte ML-modellen kunnen verder ontwikkeld worden om de nauwkeurigheid en de toepassing in de betreffende case te verbeteren. Daarbij hebben we het voornemen om de toepassing en resultaten van de drie cases voor te leggen aan een externe partij voor een peer review. Daarnaast bevelen we aan om het onderzoek naar de toepassing van ML voort te zetten met een focus op allereerst de validatie van resultaten en ten tweede het operationaliseren van de toegepaste technieken. Het regulier toepassen van deze ML-technieken door het NLR of andere partijen (denk aan bijvoorbeeld het ABL) vraagt verder onderzoek naar o.a. de praktische implementatie en instandhouding. Tenslotte raden we aan om verder onderzoek te doen naar andere ML-technieken voor de analyse van andere soorten data. Met name de toepassing van ML-technieken op radar data of vluchtdata is interessant met het oog op bijvoorbeeld het detecteren van afwijkende vluchten, het duiden van trends, en het monitoren van de effecten van genomen risicobeheersmaatregelen.

6 Referenties

1. Ministerie van Infrastructuur en Waterstaat, 2020. Nederlandse Luchtvaartveiligheidsprogramma 2020-2024. Referentie IENW/BSK-2019/256381. <https://www.rijksoverheid.nl/documenten/rapporten/2020/02/21/bijlage-1-nederlands-luchtvaartveiligheidsprogramma-2020-2024-februari-2020>
2. Verstraeten, J.G., Geest, P.J. van der, Es, G.W.H. van, Giesberts, M.K.H., Klein Obbink, B, Roelen, A.L.C., Integrale Veiligheidsanalyse Schiphol. NLR-CR-2017-313, NLR, 2018.
3. Ministerie van Infrastructuur en Waterstaat, 2019. Opzet systeemmonitor luchtvaartveiligheid. Referentie IENW/BSK-2019/251241. <https://www.rijksoverheid.nl/documenten/rapporten/2020/02/07/bijlage-8-ienw-systeemmonitor-luchtvaartveiligheid>
4. Roelen A.L.C., J.A. Post, H.H. Smit (2013), An analysis of transport airplane occurrence underreporting. NLR-CR-2013-491, NLR.
5. Cooke, R.M. (1991), Experts in Uncertainty – Opinion and Subjective Probability in Science. Environmental Ethics and Science Policy. Oxford University Press. Oxford.
6. Cooke, R.M. and L.J.H. Goosens (2000), Procedures guide for structured expert judgement, European Commission, EUR 18820 EN.
7. Odisho II, E. V., Predicting Pilot Misperception of Runway Excursion Risk Through Machine Learning Algorithms of Recorded Flight Data, Embry-Riddle Aeronautical University, 2-2020. <https://commons.erau.edu/cgi/viewcontent.cgi?article=1503&context=edt>
8. Staat van Schiphol 2019. Inspectie Leefomgeving en Transport.
9. ICAO, Artificial Intelligence (AI), [https://www.icao.int/safety/Pages/Artificial-Intelligence-\(AI\).aspx](https://www.icao.int/safety/Pages/Artificial-Intelligence-(AI).aspx)
10. Samuel, Arthur L, Some Studies in Machine Learning Using the Game of Checkers (1959), IBM Journal 535 <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.368.2254&rep=rep1&type=pdf>
11. Flach, P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, 2012.
12. ICAO Safety Management Manual (Doc 5859 4th edition, 2018), 6.5.
13. ICAO Chat with NORM, www.icao.int/safety/iStars/Pages/Chat-with-NORM.aspx (accessed 09 Nov 2020)
14. This is the information valid on 09 November 2020. Further changes can be applicable according to the search result on Global Repository of actual NOTAMS of ICAO at <https://www.icao.int/safety/iStars/Pages/notams.aspx> (accessed 09 Nov 2020)

15. Air Traffic Management, ICAO sees Artificial Intelligence as the NORM (published on 18 Sept 2018).
<https://airtrafficmanagement.keypublishing.com/2018/09/18/icao-sees-artificial-intelligence-as-the-norm/>
 (accessed 08 Nov 2020)
16. EASA, EASA AI Roadmap (February 2020),
<https://www.easa.europa.eu/sites/default/files/dfu/EASA-AI-Roadmap-v1.0.pdf>
17. ICCAIA and CANSO 'Artificial Intelligence and Digitalization in Aviation' (Working paper, 40th Session of the Assembly, A40-WP/268), dated 1 August 2019.
https://www.icao.int/Meetings/a40/Documents/WP/wp_268_en.pdf
18. Microsoft, What is a Machine Learning Model? (1 April 2019)
<https://docs.microsoft.com/en-us/windows/ai/windows-ml/what-is-a-machine-learning-model#:~:text=A%20machine%20learning%20model%20is,and%20learn%20from%20those%20data.>
19. OECD Glossary of Statistical Terms. OECD. 2008. p. 119.
20. ICAO, Annex 19 Safety Management to the Chicago Convention (2nd edition), 1-2.
21. Marcos, R. and others, Application of Machine Learning for ATM Performance Assessment – Identification Sources of En-Route Flight Inefficiency. SESAR Innovation Days 2018.
22. Demand Data Repository (DDR), restricted-access flight database maintained by EUROCONTROL and correlated position records (CPR).
23. Hegde, J. and B Rokseth, Applications of Machine Learning Methods for Engineering Risk Assessment – A Review, (2020) 122 Safety Science.
24. Rustam, F. and others, Tweets Classification on the Base of Sentiments for US Airline Companies (2019). 21 Entropy 1078.
25. Moek, G, G.J. Bakker, H.T.H. van der Zee, R. Wever Onderzoek naar anomaliedetectie in militaire vluchtdata en statistische analyse van safety logs. NLR-CR-2013-091.
26. Aviation accident database (pre 2008). (sd). Opgehaald van NTSB:
<https://www.nts.gov/ layouts/ntsb.aviation/index.aspx>
27. Cachola, I., Lo, K., Cohan, A., & S. Weld, D. (2020). TLDR: Extreme Summarization of Scientific Documents. arXiv.
28. Data.nts.gov. (sd). Opgehaald van NTSB: <https://data.nts.gov/carol-main-public/basic-search>
29. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language. CoRR.
30. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Zettlemoyer, L. (2019). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. arXiv.

31. Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. Association for Computational Linguistics, 74-81.
32. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
33. Raffel, C., Shazeer, N., Roberts, A., & Lee, K. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text TransformerExploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. arXiv.
34. Wolf, T. e. (2020). Transformers: State-of-the-Art Natural Language Processing. (pp. 38-45). Association for Computational Linguistics.
35. Nisbet, R., Elder, J., & Miner, G. (2009). Handbook of Statistical Analysis and Data Mining Applications. Academic Press.
36. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
37. National Aeronautics and Space Administration [NASA]. (2012). Sample flight data for 35 aircraft. Retrieved from the NASA DASHlink website at <https://c3.nasa.gov/DASHlink/projects/85/resources/?type=ds>
38. DuOMAn Subjectivity Lexicon (Version 1.0) (2014) [Data set]. Available at the Dutch Language Institute: <http://hdl.handle.net/10032/tm-a2-g9>

Appendix A Overzicht van databronnen met categorisering

De onderstaande tabel toont het overzicht van databronnen met bijbehorende karakteristieken zoals in paragraaf 2.2 is gedefinieerd. De volgende beknopte omschrijvingen zijn gebruikt:

Systeemmodel

- A: Internationale afspraken
- B: Regelgeving en kaders
- C: Toelating en toezicht
- D: Inrichting (vlieg)operatie
- E: Resultaat (vlieg)operatie
- F: Reflectie

Data Format

- Doc: Document
- DS: Dataset

Type eigenaar

- NL Sector: Nederlandse Luchtvaart Sectorpartij
- NL Overheid: Nederlandse Overheid
- EU Overheid: Europese overheid

Tabel A-1: Lijst met databronnen ter ondersteuning van de NVA.

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
Data_01	ABL Factsheets	Het ABL geeft periodiek factsheets uit. Hierin komen diverse onderwerpen uitgebreid aan bod die belangrijk zijn voor de luchtvaartveiligheid. Dat kunnen onderwerpen uit het dashboard zijn, maar ook bredere of juist meer specifieke onderwerpen. De actualiteit, signalen uit de sector en eigen inzicht binnen het ABL zijn hierbij leidend. Factsheets bestaan uit verder onderzoek naar de inhoud van de meldingen, het leggen van verbanden en het zoeken naar oorzaken.	Hazard	Analyse	Doc	ILT	NL Overheid	Open source Overheid	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	E
Data_02	Dashboard Luchtvaart-voorvallen	Het Analysebureau luchtvaartvoorvallen (ABL) verwerkt maandelijks het aantal gemelde voorvallen vanuit de Nederlandse luchtvaart in een dashboard. Op deze pagina staan vier categorieën uitgelicht uit het dashboard. Onderaan deze pagina staat een toelichting over de totstandkoming van dit dashboard.	Hazard	Stat	DS	ILT	NL Overheid	Proprietary Overheid	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	E
Data_03	Resultaten Audits en Inspecties	Deze audits en inspecties zijn gericht op mogelijke veiligheidsproblemen of afwijkingen ten aanzien van de geldende (Europese) wet/regelgeving.	Hazard	Analyse	Doc	ILT	NL Overheid	Proprietary Overheid	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone	C
Data_04	Signaal-rapportages ILT	Document waarmee de ILT de overheid, sector en het publiek op de hoogte stelt van (mogelijke) opkomende risico's en of gevaren.	Hazard	Analyse	Doc	ILT	NL Overheid	Open source Overheid	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone	C
Data_05	Handhavings-rapportages ILT	Een handhavingsrapportage betreft de in de Wet Luchtvaart vastgelegde veiligheids- en milieuaspecten van het luchthaven luchtverkeer. Het gaat hierbij om de bevindingen van de Inspectie Leefomgeving en	Hazard	Analyse	Doc	ILT	NL Overheid	Open source Overheid	CAT, GA	Operator, Airport, ANSP, TO, GA,	C

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
		Transport met betrekking tot de naleving van normen en regels door luchthavens, Luchtverkeersleiding Nederland en de op Nederlandse luchthavens opererende luchtvaartmaatschappijen, eventueel door de luchtvaartsector getroffen voorzieningen om normoverschrijdingen te voorkomen en eventuele maatregelen van de inspecteur-generaal.								Ground, Mx, Drone	
Data_06	Onderzoeken en kwartaal-rapportage OVV	Onderzoeken uitgevoerd door de OVV, overwegend vallend in de volgende drie categorieën: Volledig Onderzoek, Verkort Onderzoek en een vermelding in de kwartaalrapportage.	Hazard	Analyse	Doc	OVV	Overig	Open source	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	F
Data_07	Aanbevelingen expertgroep luchtvaart	Volgens kamerstuk bestaat deze groep uit bewindsleden van de Ministeries: IenW, BZ, VenJ, Defensie en BZK; KLM groep, Corendon, TUIfly en VNV.	Beleid	Mitigatie	Doc	IenW	NL Overheid	Proprietary Overheid	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	F
Data_08	Integrale Veiligheidsanalyses	Veiligheidsanalyses uitgevoerd in opdracht van het Ministerie IenW, zoals de IVA Schiphol, IVA Eindhoven en Actualisatie IVA Schiphol.	Hazard	Analyse	Doc	IenW	NL Overheid	Proprietary Overheid	CAT	Operator, Airport, ANSP, Ground	F
Data_09	Europees EPAS	The EPAS, a key component of the European Aviation Safety Programme (EASP), provides a coherent and transparent framework for safety management at regional and State level. It includes the complete set of EASA rulemaking tasks and is consistent with the goals and objectives of the ICAO Global Aviation Safety Plan (GASP).	Beleid	Mitigatie	Doc	EC	Internationale instantie	Open source	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx	A

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
Data_10	Europees EASP	This document describes how aviation safety is managed at European Union level. It was prepared by the European Commission and the European Aviation Safety Agency. It will be amended by the European Commission from time to time to reflect any changes that occur within the European aviation safety system.	Beleid	Mitigatie	Doc	EASA	Internationale instantie	Open source	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	A
Data_11	Internationaal GASP	The purpose of the Global Aviation Safety Plan (GASP) is to continually reduce fatalities, and the risk of fatalities, by guiding the development of a harmonized aviation safety strategy, and developing and implementing regional and national aviation safety plans. A safe aviation system contributes to the economic development of States and their industries. The GASP promotes the implementation of a State's safety oversight system, a risk-based approach to managing safety as well as a coordinated approach to collaboration between States, regions and industry. States are encouraged to support and implement the GASP as the strategy for the continuous improvement of global aviation safety	Beleid	Mitigatie	Doc	ICAO	Internationale instantie	Open source	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	A
Data_12	Jaarverslagen luchthavens	Jaarverslagen bevatten vaak safety performance indicators zoals birdstrikes en runway incursions, en ook andere ontwikkelingen op het gebied van veiligheid.	Hazard	Analyse	Doc	Sector partij	NL Sector	Open source	CAT, GA	Airport	D
Data_13	EASA Annual Safety Review	Jaarlijkse EASA rapportage waarin een analyse van occurrences plaatsvindt om een totaalbeeld te schetsen van de luchtvaartveiligheid van Europa van het afgelopen jaar.	Hazard	Analyse	Doc	EASA	Internationale instantie	Open source	CAT, GA, Drone	Operator, Airport, ANSP, GA, Ground, Drone, OEM	E
Data_14	METAR	Weerinformatie zoals wind (kracht, richting), neerslag, mist, etc.	Context	Ruwe data	DS	KNMI	NL Overheid	Open source Overheid	CAT, GA, Drone	Operator, Airport	D

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
Data_15	ARTAS Radar tracks	Landelijk Surveillance radar netwerk in Nederland. Hoeveelheid verkeer, type verkeer, vliegbewegingen per luchthaven, overvliegers.	Context	Ruwe data	DS	LVNL	NL Sector	Proprietary	CAT, GA	Operator, Airport, ANSP, TO, GA, Drone	D
Data_16	ADS-B Mode S tracks	Wereldwijde beschikbaarheid van ADS-B mode S data. Hoeveelheid verkeer, type verkeer, vliegbewegingen per luchthaven, overvliegers.	Context	Ruwe data	DS	Meerdere	Overig	Proprietary	CAT	Operator, Airport, ANSP	D
Data_17	Grondradar Luchthavens	Hoeveelheid verkeer, type verkeer op de grond.	Context	Ruwe data	DS	Meerdere	NL Sector	Proprietary	CAT, GA	Operator, GA, Ground	D
Data_18	Gebruikte baan-combinatie (Schiphol)	Start- en vertrekrichting.	Context	Ruwe data	DS	LVNL	NL Sector	Proprietary	CAT	Operator, Airport, ANSP	D
Data_19	Baankruisingen (Schiphol)	Informatie over baankruisingen (runway crossings).	Context	Ruwe data	DS	LVNL	NL Sector	Proprietary	CAT	Operator, Airport, ANSP	D
Data_20	Late baan-wisselingen	Informatie over baanwisselingen (runway configuratie), (on)voorspelbaarheid baangebruik.	Context	Ruwe data	DS	LVNL	NL Sector	Proprietary	CAT	Operator, Airport, ANSP	D
Data_21	Fanomos Radar Tracks	Hoeveelheid verkeer, type verkeer, vliegbewegingen per luchthaven, overvliegers.	Context	Ruwe data	DS	NLR	NL Sector	Proprietary	CAT, GA	Operator, TO, GA	D
Data_22	OAG Time schedule	Vertrektijden vluchten, hoeveelheid vluchten, type kisten.	Context	Ruwe data	DS	OAG	Overig	Proprietary	CAT	Operator, Airport	D
Data_23	Voorval meldingen aan ABL	Het Analysebureau luchtvaartvoorvallen (ABL) verzamelt de meldingen van voorvallen, om vroegtijdig trends te signaleren. Het ABL deelt de inzichten die uit de analyses komen met de luchtvaartsector zodat deze eventueel maatregelen kan nemen om de vliegveiligheid te verbeteren. Het betreft hier de meldingen zelf als databron.	Hazard	Ruwe data	DS	ILT	NL Overheid	Proprietary Overheid	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone	E
Data_24	Voorval meldingen Sector	Voorvallen	Hazard	Ruwe data	DS	Meerdere	NL Sector	Proprietary	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx,	E

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
										Drone, OEM	
Data_25	FDM Events	Vluchtafwijkingen	Context	Ruwe data	DS	Meerdere	NL Sector	Proprietary	CAT, GA	Operator, TO, GA	D
Data_26	Runway Meteo	Weerinformatie zoals weer(kracht, richting), neerslag, mist, etc.	Context	Ruwe data	DS	LVNL	NL Sector	Open source	CAT, GA	Operator, TO, GA	D
Data_27	Staat van Schiphol	De Inspectie Leefomgeving en Transport (ILT) brengt jaarlijks de Staat van Schiphol uit waarmee inzicht wordt geboden in de ontwikkeling van veiligheid, duurzaamheid en leefomgeving op en rondom Schiphol. De Staat geeft naast algemene feitelijke informatie ook breed en actueel inzicht in de thema's 'veilig vliegen', 'veilig en gezond leven' en 'veilig en gezond werken'. Op basis van de Staat concludeert de ILT of normen worden overschreden of dat risico's toe- of afnemen.	Hazard	Analyse	Doc	ILT	NL Overheid	Open source Overheid	CAT	Operator, Airport, GA, Drone	E
Data_28	Landelijke Systeem Monitor	De systeemmonitor geeft informatie over het functioneren van het luchtvaartveiligheids-systeem. De minister van IenW biedt de systeemmonitor periodiek aan de Tweede Kamer aan.	Beleid	Stat	Doc	IenW	NL Overheid	Open source Overheid	CAT, GA	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	F
Data_29	Luchtvaartuig Register	Aantal geregistreerde vliegtuigen, type eigenaar (privaat, vliegschool, luchtvaart-maatschappij).	Context	Ruwe data	DS	ILT	NL Overheid	Open source Overheid	CAT, GA, Drone	Operator, TO, GA, Drone	D
Data_30	NLVP	Het Nederlands Luchtvaartveiligheids-programma (NLVP) beschrijft hoe de veiligheid van de burgerluchtvaart binnen Nederland is geborgd in samenhang tussen beleid, toezicht en de luchtvaartorganisaties.	Beleid	Analyse	Doc	IenW	NL Overheid	Open source Overheid	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	B

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
Data_31	NALV	Met het Nederlandse actieplan voor de luchtvaart-veiligheid (NALV) geeft Ministerie IenW uitvoering aan het NLVP.	Beleid	Mitigatie	Doc	IenW	NL Overheid	Open source Overheid	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	C
Data_32	Social Media	Brede contextuele informatie over vluchten (van spotters, (hobby-)piloten, burgers), social media posts over bijvoorbeeld drone-gebruik, gemaakte (hobby) vluchten, incidenten.	Context	Ruwe data	Doc	Publiek	Overig	Open source	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	D
Data_33	Radio-communicatie	Informatie over mogelijke vluchtafwijkingen, incidenten.	Context	Ruwe data	DS	Spotters	Overig	Proprietary	CAT, GA, Drone	Operator, TO, GA, Drone	D
Data_34	ANS Performance ³	Informatie over luchtruim complexiteit, prestaties van luchtverkeersleiding, financiële prestaties.	Context	Stat	DS	Eurocontrol	EU Overheid	Open source	CAT	Operator, Airport, ANSP	E
Data_35	EU Performance Scheme (Reference Periods) ⁴	Safety culture, financiële prestaties, SMS effectiviteit (effectiveness of safety management, safety culture).	Context	Stat	DS	Eurocontrol	EU Overheid	Open source	CAT	Airport, ANSP	E
Data_36	EVAIR Safety Report ⁵	Through the EVAIR reporting process, Eurocontrol gathers data about and analyse a wide spectrum of safety concerns, incidents and accidents. The topics we see most frequently and focus on as part of our bulletin include runway incursion, level-bust, call-sign similarities, loss of communication and drones.	Hazard	Analyse	Doc	Eurocontrol	EU Overheid	Open source	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	E

³ <https://ansperformance.eu/data/>

⁴ <https://www.eurocontrol.int/prudata/dashboard/vis/2017/>

⁵ <https://www.eurocontrol.int/publication/eurocontrol-voluntary-atm-incident-reporting-evair-safety-bulletin-20>

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
Data_37	Performance Review Report ⁶	The Performance Review Report reviews the performance of air traffic management in Europe per year. It looks at the following key performance areas: safety, capacity, cost-effectiveness and environment.	Context	Stat	Doc	Eurocontrol	EU Overheid	Open source	CAT, GA	Airport, ANSP	E
Data_38	Network of Analysts (NoA)	The NoA comprises EASA, European Commission, EASA Members States and an observer. It was initially formed as a voluntary network to support the analysis of safety data for the European Aviation Safety Plan, which is the predecessor to the EPAS, in 2011. The implementation of EU Regulation 376/2014 has formalised the role of the NoA and underscored the importance of safety analysis in supporting the EPAS and improvement of aviation safety in Europe. The NoA currently meets twice a year, once in Spring and once in Autumn.	Hazard	Analyse	Doc	EASA	EU Overheid	Proprietary	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	F
Data_39	Collaborative Analysis Groups (CAGs)	The CAGs are expert groups, responsible for analysing the safety of European aviation. Each CAG works on a domain and its membership is derived from key stakeholders in the domain.	Hazard	Analyse	Doc	EASA	EU Overheid	Proprietary	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	F
Data_40	Dossier luchtvaart	Diverse datasets van het CBS over luchtvaart.	Context	Stat	DS	CBS	Overig	Open source	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	E

⁶ <https://www.eurocontrol.int/sites/default/files/2019-06/pr-2018.pdf>

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
Data_41	Externe veiligheidsanalyses	Veiligheidsanalyses met doel om de veiligheid van een onderdeel van de luchtvaart te vergroten. Vaak uitgevoerd door een onderzoeksbureau in opdracht van een sector partij.	Hazard	Mitigatie	Doc	Onderzoeksbureaus	Overig	Proprietary	CAT, GA	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	E
Data_42	Toezietsprogramma (Schiphol/ Luchtvaart)	Dit programma levert een bijdrage aan een veilig en duurzaam Schiphol. De belangrijkste vraag is: wat zijn de risico's en wat is de samenhang tussen de risicothema's? Het gaat om risico's op het gebied van vliegveiligheid, maar ook van arbeidsveiligheid, hinderbeleving, gezondheid en leefbaarheid.	Beleid	Analyse	Doc	ILT	NL Overheid	Open source Overheid	CAT	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	C
Data_43	Top-5 ground / flight safety hazards / risks	De top 5 van risks voor ground en air zoals gedefinieerd door het ISMS en de aangesloten sector partijen.	Hazard	Analyse	DS	ISMS	NL Sector	Proprietary	CAT	Operator, Airport, ANSP, Ground, Drone, OEM	E
Data_44	ISMS Roadmap	De uit het ISMS voortvloeiende maatregelen voor een verbeterde veiligheid vormen samen de Roadmap Veiligheidsverbetering Schiphol, waarvan een samenvatting op deze website staat. We laten ook zien hoe we opvolging geven aan de aanbevelingen van de Onderzoeksraad voor Veiligheid en een veiligheidsanalyse die is uitgevoerd door het Nederlands Lucht- en Ruimtevaartcentrum (NLR).	Hazard	Mitigatie	Doc	ISMS	NL Sector	Open source	CAT	Operator, Airport, ANSP, Ground, Drone, OEM	D
Data_45	Evaluatie ISMS Maatregelen	Externe evaluaties van ISMS Roadmap maatregelen. Meestal uitgevoerd door een externe partij zoals het ov.	Hazard	Analyse	Doc	ISMS	NL Sector	Proprietary	CAT	Operator, Airport, ANSP, Ground, Drone, OEM	C

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
Data_46	Externe evaluatie ISMS	Jaarlijkse externe evaluatie van het ISMS. In 2019 vond de eerste evaluatie plaats, uitgevoerd door Baines and Simmons.	Hazard	Analyse	Doc	Onderzoeks-bureaus	Overig	Open source	CAT	Operator, Airport, ANSP, Ground	D
Data_47	Externe evaluatie ABL	Jaarlijkse externe evaluatie van het ABL. In 2019 vond de eerste evaluatie plaats, uitgevoerd door Baines and Simmons.	Hazard	Analyse	Doc	Onderzoeks-bureaus	Overig	Open source	CAT	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone, OEM	C
Data_48	Rijksbegroting	Rijksbegroting die wordt gepubliceerd op Prinsjesdag. Geeft inzicht in welke budgetten er beschikbaar zijn en hoe deze zich verhouden tot de budgetten van voorgaande jaren. Informatie over beschikbare budget, trends, bezuinigingen in luchtvaartveiligheid (toezicht, beleid), FTE DGLM, ILT luchtvaart (gepland vs gerealiseerd).	Context	Stat	Doc	Ministerie van Financien	NL Overheid	Open source Overheid	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone	B
Data_49	Meeting notes ABL Meetings CA+	Aantal meetings en aanwezigheid, informatie over effectiviteit ABL CA+.	Hazard	Analyse	Doc	ILT	NL Overheid	Proprietary Overheid	CAT	Operator, Airport, ANSP, OEM	C
Data_50	Meeting notes ABL Meetings GA+	Aantal meetings en aanwezigheid, informatie over effectiviteit ABL GA+.	Hazard	Analyse	Doc	ILT	NL Overheid	Proprietary Overheid	GA	Airport, ANSP, TO, GA, Drone, OEM	C
Data_51	Meeting notes ISMS TOPSAG	Aantal meetings en aanwezigheid, informatie over effectiviteit ISMS.	Hazard	Analyse	Doc	ISMS	NL Sector	Proprietary	CAT	Operator, Airport, ANSP, Ground, Drone, OEM	E

Data_n	Databron	Beschrijving	Informatie	Analyse niveau	Data Format	Eigenaar	Type eigenaar	Toegankelijkheid	Domein	Service provider	Systeem model
Data_52	LVNL Overzicht Voorvallen	LVNL publiceert op deze pagina proactief de melding van potentieel ernstige voorvallen en de resultaten van het voorvalonderzoeken. Op deze manier krijgt de omgeving meer inzicht in vliegveiligheid. Per voorval zijn de onderzoeksresultaten te lezen en geeft LVNL aan welke maatregelen LVNL neemt om voorvallen te voorkomen.	Hazard	Analyse	Doc	LVNL	NL Sector	Open source	CAT	Operator, Airport, ANSP, Ground, Drone	E
Data_53	Nederland wereldwijd reisadviezen	Conflictgebieden die risico's vormen voor Nederlandse luchtvaart.	Context	Analyse	Doc	Ministerie van buitenlandse zaken	NL Overheid	Open source Overheid	CAT	Operator, GA	D
Data_54	ILT Jaarverslag	Terugblik op het voorgaande jaar vanuit ILT oogpunt. Jaarverslag bevat doorgaans interessante statistiek met betrekking tot de organisatie en werkzaamheden van de ILT. Detailniveau verschilt per verslag.	Context	Stat	Doc	ILT	NL Overheid	Open source Overheid	CAT, GA, Drone	Operator, Airport, ANSP, Ground, Drone, OEM	C
Data_55	SMM	Safety management system (SMS) manual.	Beleid	Ruwe data	Doc	Sector partij	NL Sector	Proprietary	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone	D
Data_56	OM	Operations manual.	Beleid	Ruwe data	Doc	Sector partij	NL Sector	Proprietary	CAT, GA, Drone	Operator, Airport, ANSP, TO, GA, Ground, Mx, Drone	D
Data_57	ICAO Emerging Issues and Ops Safety Risk informatie	ICAO Emerging Issues and Ops Safety Risk informatie (te verkrijgen via ICAO's Secured Portal) met informatie over: High Risk Categories of Occurrences (HRC); Significant Safety Concerns (SSC); Safety Oversight Index (SOI) score (overall en per gebied (3))	Hazard	Stat	DS	ICAO	Overig	Proprietary	CAT	Operator, Airport, ANSP, Ground, Mx	E

Databronnen die informatie leveren over statistieken, analyses en mitigaties bevatten meestal informatie die gebaseerd is op (onderliggende) ruwe data. Databronnen die aangemerkt zijn met analyseniveau “statistiek”, “analyse” en “mitigatie” zijn gegroepeerd in een aantal thema’s: Beleidseffectiviteit, Beleidstukken, Toezichtsrapportages (ILT), ISMS Roadmap, Overlegorgaan, SMS effectiviteit, Veiligheidsanalyses en veiligheidsprestatie, voorvalonderzoeken.

Tabel A-2: Indeling van databronnen naar een groepering voor het vaststellen van databronnen onderliggende aan diverse type informatie (onder groepering).

Data_n	Databron	Groepering
Data_01	ABL Factsheets	Voorvalonderzoeken
Data_02	Dashboard Luchtvaartvoorvallen	Voorvalonderzoeken
Data_03	Resultaten Audits en Inspecties	Toezichtsrapportages ILT
Data_04	Signaalrapportages ILT	Toezichtsrapportages ILT
Data_05	Handhavingsrapportages ILT	Toezichtsrapportages ILT
Data_06	Onderzoeken en kwartaalrapportage OVV	Veiligheidsanalyses en veiligheidsprestatie
Data_07	Aanbevelingen expertgroep luchtvaart	Overlegorgaan
Data_08	Integrale Veiligheidsanalyses	Veiligheidsanalyses en veiligheidsprestatie
Data_09	Europees EPAS	Beleidstukken
Data_10	Europees EASP	Beleidstukken
Data_11	Internationaal GASP	Beleidstukken
Data_12	Jaar verslagen luchthavens	Beleidseffectiviteit
Data_13	EASA Annual Safety Review	Voorvalonderzoeken
Data_27	Staat van Schiphol	Veiligheidsanalyses en veiligheidsprestatie
Data_28	Landelijke Systeem Monitor	Beleidseffectiviteit
Data_30	NLVP	Beleidstukken
Data_31	NALV	Beleidstukken
Data_34	ANS Performance	Diverse datasets
Data_35	EU Performance Scheme (Reference Periods)	Beleidseffectiviteit
Data_36	EVAIR Safety Report	Voorvalonderzoeken
Data_37	Performance Review Report	Veiligheidsanalyses en veiligheidsprestatie
Data_38	Network of Analysts (NoA)	Overlegorgaan
Data_39	Collaborative Analysis Groups (CAGs)	Overlegorgaan
Data_40	Dossier luchtvaart	Diverse datasets
Data_41	Externe veiligheidsanalyses	Veiligheidsanalyses en veiligheidsprestatie
Data_42	Toezichtsprogramma (Schiphol/Luchtvaart)	Beleidseffectiviteit
Data_43	Top-5 ground/flight safety hazards/risks	Veiligheidsanalyses en veiligheidsprestatie
Data_44	ISMS Roadmap	ISMS Roadmap
Data_45	Evaluatie ISMS Maatregelen	ISMS Roadmap
Data_46	Externe evaluatie ISMS	SMS effectiviteit
Data_47	Externe evaluatie ABL	SMS effectiviteit
Data_48	Rijksbegroting	Diverse datasets
Data_49	Meeting notes ABL Meetings CA+	Overlegorgaan
Data_50	Meeting notes ABL Meetings GA+	Overlegorgaan
Data_51	Meeting notes ISMS TOPSAG	Overlegorgaan
Data_52	LVNL Overzicht Voorvallen	Voorvalonderzoeken
Data_53	Nederland wereldwijd reisadviezen	Diverse datasets
Data_54	ILT Jaarverslag	Beleidseffectiviteit
Data_57	ICAO Emerging Issues and Ops Safety Risk informatie	Veiligheidsanalyses en veiligheidsprestatie

Appendix B Lijst met voorbeelden van contextuele data

In dit onderzoek hebben we op basis van beschikbare kennis, ervaring en eerdere veiligheidsanalyses een lijst opgesteld met voorbeelden van contextuele data die mogelijk nuttig of nodig zijn in een NVA. Deze contextuele data kunnen gebruikt worden voor verrijking en normalisatie van veiligheids- en operationele data. Er zijn twee soorten contextuele data te onderscheiden, namelijk data over het activiteitsniveau of de geleverde ‘productie’ per systeemelement (zie het model in paragraaf 1.4), en ten tweede data met achtergrondinformatie over het ‘systeem’. Tabel B-1 geeft een overzicht van mogelijke contextuele data. De tabel bevat de volgende informatie in de kolommen:

- Contextdata_n: Identificatienummer van de contextuele data.
- Data_n: Identificatienummer van de databron die deze contextuele data bevat (zie ook Tabel A-1, kolom “informatie”). Er zijn aantal soorten contextinformatie geïdentificeerd waar we nog geen passende databron voor hebben geïdentificeerd. Het is onduidelijk of deze informatie wel of niet verzameld wordt, en indien wel, in welke databron deze informatie beschikbaar is
- Onderwerp: Beknopte beschrijving van de contextuele data.
- Type: De referentie A t/m F verwijzen naar de elementen uit het systeemmodel; de referentie ‘systeem’ verwijst naar contextuele informatie over het (luchtvaart)systeem.

Tabel B-1: Lijst met mogelijke contextdata voor verrijkingen en normalisaties.

Contextdata_n	Data_n	Onderwerp	Type
Contextdata_1	Data_10, Data_11, Data_30	Beleidsdoelstellingen nationaal en internationaal EI% implementatie	A: Internationale afspraken
Contextdata_2	Data_42, Data_48	Budget DGLM	B: Regelgeving en kaders
Contextdata_3	Data_42, Data_48	Budget ILT	B: Regelgeving en kaders
Contextdata_4	Data_42, Data_54	Aantal inspecteurs (FTE's) ILT luchtvaart	B: Regelgeving en kaders
Contextdata_5	Data_42, Data_48	Aantal medewerkers (FTE's) DGLM	B: Regelgeving en kaders
Contextdata_6	Data_30, Data_31	NLVP en NALV Publicatie status	B: Regelgeving en kaders
Contextdata_7	Data_30, Data_31	NLVP en NALV Evaluatie en update status	B: Regelgeving en kaders
Contextdata_8	Data_05	Aantal vergunningaanvragen	C: Toelating en toezicht
Contextdata_9	Data_05	Aantal afgewezen aanvragen	C: Toelating en toezicht
Contextdata_10	Data_05	Aantal toegekende aanvragen	C: Toelating en toezicht
Contextdata_11	Data_03	Aantal ingetrokken vergunningen	C: Toelating en toezicht
Contextdata_12	-	Aantal safety cases beoordeeld	C: Toelating en toezicht
Contextdata_13	Data_05	Aantal licenties en type licenties verstrekt	C: Toelating en toezicht
Contextdata_14	Data_01, Data_02, Data_23, Data_24, Data_47, Data_49, Data_51	Effectiviteit ABL bijv. compleetheid van meldingen (aantal), bruikbaarheid van meldingen (inhoud), aantal verbetermaatregelen n.a.v. meldingen	C: Toelating en toezicht
Contextdata_15	-	Aantal geplande inspecties	C: Toelating en toezicht
Contextdata_16	Data_54	Aantal uitgevoerde inspecties	C: Toelating en toezicht
Contextdata_17	Data_54	Aantal uitgedeelde interventies	C: Toelating en toezicht
Contextdata_18	Data_03, Data_04, Data_05, Data_41	Aantal signaalrapportages	C: Toelating en toezicht
Contextdata_19	Data_03, Data_04, Data_05, Data_41	Aantal uitgedeelde processen-verbaal (luchtvaartpolitie/ILT)	C: Toelating en toezicht
Contextdata_20	Data_03, Data_04, Data_05, Data_41	Aantal onder toezicht staande bedrijven	C: Toelating en toezicht
Contextdata_21	Data_03, Data_04, Data_05, Data_41	Aantal onder verscherpt toezicht staande bedrijven	C: Toelating en toezicht

Contextdata_n	Data_n	Onderwerp	Type
Contextdata_22	-	Effectiviteit en volwassenheid ISMS	D: Inrichting (vlieg)operatie
Contextdata_23	Data_03, Data_04, Data_05, Data_43	Effectiviteit en volwassenheid SMS sectorpartijen	D: Inrichting (vlieg)operatie
Contextdata_24	Data_03, Data_04, Data_05, Data_12	Effectiviteit en volwassenheid veiligheidscultuur sectorpartijen	D: Inrichting (vlieg)operatie
Contextdata_25	Data_03, Data_04, Data_05, Data_12	Effectiviteit en volwassenheid occurrence reporting systeem (verplicht, vrijwillig) sectorpartijen	D: Inrichting (vlieg)operatie
Contextdata_26	Data_12, Data_13, Data_23, Data_24, Data_25, Data_29, Data_36, Data_52	Genormaliseerd aantal incidenten per type service provider, per CICTT occurrence category, flight phase, per etc.	E: Resultaat (vlieg)operatie
Contextdata_27	Data_43, Data_44, Data_45, Data_46, Data_56	Status top 5 ISMS risico's Ground	E: Resultaat (vlieg)operatie
Contextdata_28	Data_43, Data_44, Data_45, Data_46	Status top 5 ISMS risico's Air	E: Resultaat (vlieg)operatie
Contextdata_29	-	Aantal expertsessies luchtvaartveiligheid	F: Reflectie
Contextdata_30	-	Aantal gevoerde overleggen en safety promotiecampagnes	F: Reflectie
Contextdata_31	Data_06	Aantal uitgevoerde onderzoeken – volledig/verkort/niet onderzocht (OVV)	F: Reflectie
Contextdata_32	Data_06	Aantal gepubliceerde aanbevelingen (OVV)	F: Reflectie
Contextdata_33	Data_06, Data_28	Aantal opgevolgde/afgesloten, niet uitgevoerde en openstaande aanbevelingen (OVV)	F: Reflectie
Contextdata_34	Data_06, Data_28	Duur van openstaande aanbevelingen (OVV)	F: Reflectie
Contextdata_35	Data_02	Aantal dashboards verversingen per jaar (ABL)	F: Reflectie
Contextdata_36	Data_01	Aantal gepubliceerde ABL factsheets per jaar (ABL)	F: Reflectie
Contextdata_37	-	Aantal themastudies per jaar (ABL en OVV)	F: Reflectie
Contextdata_38	Data_02	Aantal dashboards met SPIs (ABL)	F: Reflectie
Contextdata_39	Data_49, Data_50	Aantal ABL CA+ / GA+ overleggen per jaar	F: Reflectie
Contextdata_40	Data_23, Data_24	Aantal initiële meldingen per domein CAT, GA, drone, per tijdseenheid, per type operatie, etc. (ABL en sector)	F: Reflectie
Contextdata_41	-	Aantal meldingen met follow-up per domein CA, GA, drone categorie, etc. (ABL)	F: Reflectie
Contextdata_42	-	Mate van underreporting aan ABL (ABL)	F: Reflectie
Contextdata_43	-	Aantal voorlichtingsbijeenkomsten m.b.t. EU Reg 2014/376	F: Reflectie
Contextdata_44	Data_08, Data_43, Data_44, Data_45, Data_46, Data_51	Effectiviteit ISMS Maatregelen	F: Reflectie
Contextdata_45	-	Status organiseren nationale jaarlijkse veiligheidsnetwerk workshops	F: Reflectie
Contextdata_46	Data_02	Aantal updates per jaar van dashboards met SPIs (ABL)	F: Reflectie
Contextdata_47	-	Aantal aanwezigen op ABL CA+ en GA+ meetings	F: Reflectie
Contextdata_48	Data_15, Data_16, Data_17, Data_21, Data_22, Data_27	Aantal starts/landingen/bewegingen per CAT, GA, drone, per tijdseenheid, per etc.	Systeem

Contextdata_n	Data_n	Onderwerp	Type
Contextdata_49	-	Aantal vliegreun, per CAT, GA, drone, per tijdseenheid, per etc.	Systeem
Contextdata_50	Data_15, Data_16, Data_21	Aantal IFR en VFR flights, per tijdseenheid, per etc.	Systeem
Contextdata_51	Data_15, Data_16, Data_21, Data_22, Data_53	Aantal vluchten over/rondom conflictgebieden en/of naar conflict landen	Systeem
Contextdata_52	Data_19	Aantal baankruisingen per tijdseenheid, per etc.	Systeem
Contextdata_53	Data_18 Data_20	Aantal baanwisselingen per tijdseenheid, per etc.	Systeem
Contextdata_54	Data_29	Aantal vliegtuigen (CAT, GA, drone) in luchtvaartregister	Systeem
Contextdata_55	Data_40	Aantal vervoerde passagiers	Systeem
Contextdata_56	Data_40	Aantal tonnen vervoerde vracht	Systeem
Contextdata_57	Data_15, Data_16, Data_21, Data_22	Aantal dag/nacht landingen	Systeem
Contextdata_58	Data_17	Aantal sleepbewegingen	Systeem
Contextdata_59	-	Aantal FOD inspecties, bird control, runway inspecties	Systeem
Contextdata_60	Data_34, Data_35, Data_37	(Gemiddelde) vertraging	Systeem
Contextdata_61	Data_34, Data_35, Data_37	(Gemiddelde) punctualiteit	Systeem
Contextdata_62	Data_15, Data_16, Data_21	(Gemiddelde) aantal bestemmingen	Systeem
Contextdata_63	-	Aantal trainingsorganisaties	Systeem
Contextdata_64	-	Aantal vliegclubs	Systeem
Contextdata_65	-	Aantal vliegclub leden	Systeem
Contextdata_66	-	Aantal en duur storingen van infrastructuur (ANSP, luchthaven)	Systeem
Contextdata_67	-	Aantal afhandelaren	Systeem
Contextdata_68	-	Aantal afhandelingen per tijdseenheid	Systeem
Contextdata_69	-	Aantal keren dat passagiers op remote stand werden afgehandeld en aantal keren via gate/terminal	Systeem
Contextdata_70	Data_14, Data_26	Weerscondities (wind, bewolking, zichtcondities, neerslag, adverse weather).	Systeem
Contextdata_71	Data_14, Data_26	Baancondities (nat, droog, etc.)	Systeem
Contextdata_72	Data_15, Data_16, Data_21, Data_29	Diversiteit van vliegverkeer, gepresenteerd d.m.v. vloot mix, d.w.z. samenstelling vliegtuig en helikoptertypes, verdeling types over propulsion types, mass group, type of operation, type of operator, etc.	Systeem
Contextdata_73	Data_29	Gemiddelde leeftijd vloot vliegtuigen en helikopters	Systeem
Contextdata_74	Data_29	Verdeling vloot over generaties	Systeem
Contextdata_75	-	Maat voor ruimtelijke ordening, bijvoorbeeld aantal windmolenparken en andere obstakels in nabijheid vliegvelden	Systeem
Contextdata_76	Data_32	Sentiment m.b.t. luchtvaartbeleid, luchtvaart onderwerpen (thema's) etc.	Systeem

Appendix C Literatuuronderzoek toepassing Machine Learning

Appendix C.1 Introduction

One task of this project studies the use of Machine Learning (ML) techniques for a potential application in the State Safety Analysis, the Nationale Veiligheidsanalyse (NVA). The literature research has two objectives. Firstly, it will explore various ML techniques which may be useful to support the NVA. Secondly, it will identify interesting and relevant applications of ML by presenting use cases within and outside the scope of aviation. As an introduction, this appendix will briefly discuss ML to provide background information on this topic. Specifically, various ML subfields and techniques are defined and described. The introduction will be followed by several case studies which utilize ML techniques in the aviation domain. These case studies were selected considering their potential for a similar application in the NVA.

Appendix C.2 Introduction to Machine Learning

Machine Learning (ML) is known as a field of expertise of the Artificial Intelligence (AI) spectrum, which has gained much attention in the past years [9]. There is no single definition of ML used world-wide. In 1959 already, one reference defined it as ‘field of study that gives computers the ability to learn without being specifically programmed’ [10]. First self-learning computer programs date from 1959 and comprised of computers being able to learn how to play a game of checkers. In 2012, another study defined ML as the systematic study of algorithms and systems that improve their knowledge or performance with experience [11]. One of the most iconic examples of ML practices is email spam recognition. With help of a classification model, the service provider filters spam e-mails from the others, on the basis of the contents of the email. This model is usually trained on spam emails, which were received previously by other email clients or even in your own account. In other words, based what a user classifies as spam email, the e-mail client will adjust its future predictions.

This research study takes the definitions of AI and ML from the European Aviation Safety Agency (EASA) due to the lack of definitions provided by ICAO. EASA tried to define the terms in its roadmap on the use of AI as follows [16]:

- AI is ‘any technology that appears to emulate the performance of a human’; and
- ML means ‘algorithms whose performance improves as they are exposed to data’.

The definitions of EASA contain the essence of both technologies, which are widely defined in the field of data science. EASA categorizes ML into three different types: supervised, unsupervised and reinforcement methods. Each of these three different types are described in greater detail in the next sections. The importance of the AI and ML is currently underlined by ICAO and EASA, and many other national governments [17].

With the advent of big data various fields of industry recognize the usefulness of AI due to the fact that AI and ML can efficiently process a large amount of data. This capability enables various applications in the aviation sector. International Civil Aviation Organization (ICAO) supports the use of AI within the aviation industry. Mainly the domains of operational safety management, finance, and maintenance repair and overhaul (MRO) create large datasets. The most recent editions of ICAO publications show the interest of the aviation industry in the use of data, emphasizing

the importance of data-driven decision-making (D3M) [12, 13]. ICAO also promotes the use of AI in the field of aeronautical information management. The Notices to Airmen (NOTAM) Organizational and Recognition Model (NORM) [13] is an example. A NOTAM contains information which is critical to the pilot in the course of operation of an aircraft. Using the NOTAM, pilots can obtain information on potential hazards along the flight route or at a location that could affect the safety of the flight. However, it is difficult for pilots to focus more on rather critical information in the pool of NOTAMS each year. In the Netherlands alone, 236 NOTAMS are found to be published, which makes part of 12.714 NOTAMS in the European region in 2020 [14]. The Flight Service Bureau of ICAO collected more than 17.000 responses from air traffic controllers, dispatchers and pilots in the scoring method of the criticality of NOTAMS to trained NORM. According to an interview with Marco Merens, chief of integrated aviation analysis of ICAO, 'NORM can understand the context and classifies criticality scores from 1 to 5, which represents 'not significant' to 'very critical' [15].

Appendix C.3 Types of Machine Learning algorithms

ML is a computer science domain about modelling and training a computer program to learn from data and reason an output [18]. Depending on the presence of supervision and logic, ML can be categorized as supervised and unsupervised. Between a supervised and an unsupervised method, there are semi-supervised learning and reinforcement algorithms.

A supervised learning algorithm is 'the process of learning a function that maps an input to an output based on example input-output training samples.' The ML-algorithm is trained by the examples with a goal to differentiate what falls outside the training data set. For instance, a training data set for supervised ML can be combination of a set of pictures of cats and dog and names of animals as an input. The output data for the training would be that identification of animals. With this training, the model will be able to identify corresponding names (label) of the animals in the pictures. However, the model would not be able to accurately guess a name of an animal which was not provided as a training data set, but do its best to logically guess. For instance, when there are pictures of tigers, the model would guess that the picture is a cat, with a probability of for example 60%.

On the other hand, an unsupervised learning algorithm can be used 'where there is no labelled data set available to learn from'. With a data set without labels is provided, the neural network analyses the data set to find a logic itself. The predictions of the unsupervised learning model are purely based on the dataset that is given to it at that time. Predictions are not based on data that the model might have seen before.

A semi-supervised learning algorithm requires some labelled training data to be present, however training data is primarily needed for the predictions that are of interest. For example, a semi-supervised model that is able to recognize if a certain news article falls within the sports section. The training data would only include sports articles. This model will subsequently not be able to predict any of the other sections like economics, lifestyle or politics. It can only predict whether a certain article falls within the sports category or not. This model is semi-supervised, since it did not require training data for the other article categories.

A reinforcement algorithm is 'the neural network is reinforced for positive results, and punished for a negative result, forcing the neural network to learn over time'. The model is forced to learn constantly. The developer force the program by rewarding the accuracy of the results. This requires constant interaction between the developer and learning model, either by hand or by means of automatic scoring.

The image below represents types of ML mentioned above, and methods falling under each type. The category under which some methods fall is sometimes inconclusive. For example, in the image below classification is categorized as being part of the semi-supervised learning spectrum. However, there are also forms of classification that are solely part of the supervised learning spectrum.

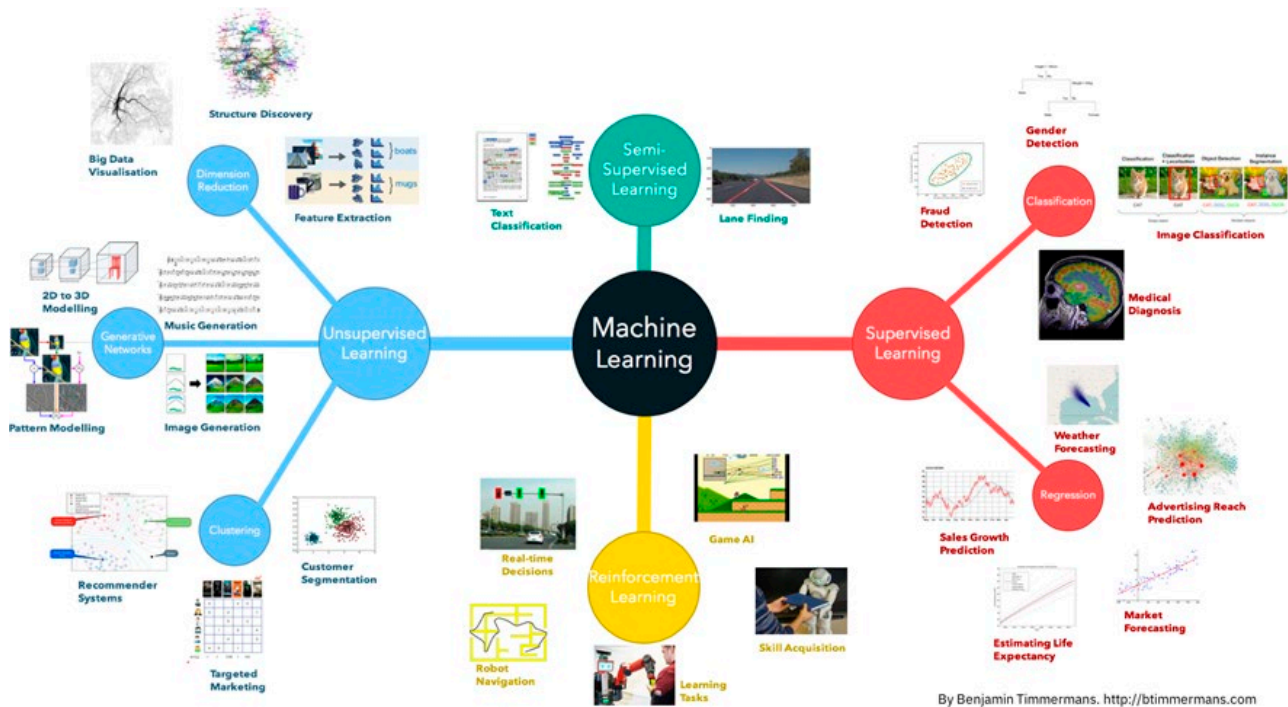


Figure C-1: Different types of Machine Learning algorithms.

Appendix C.4 Data for ML application

Data are characteristics or information, usually numerical, that are collected through observations [19]. Any type of data which can be used for the enhancement of safety is within scope of ML applications, from twitter posts to the number of passengers per flight. ICAO categorizes some data specifically under the title of 'safety data'. Safety data is 'a defined set of facts or set of safety values collected from various aviation-related sources, which is used to maintain or improve safety'. Such data is collected from proactive or reactive safety-related activities and include accident or incident investigations, safety reporting, continuing airworthiness reporting, operational performance monitoring, inspections, audits, surveys, and safety studies and reviews. [20]

Besides safety data, an important data source for ML applications is what is referred to as 'operational data' or contextual data. Examples include flight data, radar tracks, weather data, flight plan data. These data do not directly provide information on hazards or safety performance, but need to be integrated and analyzed in relation to safety data to extract meaningful information for aviation safety. As an example, flight data falls under the category of operational data. However, once it is processed and integrated with weather data, it can become 'safety data', which provides information for example of the relation between unstable approaches and wind conditions.

Insightful and valuable ML models depend on the quality and quantity of the data available. When it comes to quantity, the amount of data should be sufficient. As the word 'sufficient' may suggest, the adequate amount of data

required depends on the research problem at hand. Therefore, the sufficiency varies. However, it is still common that the required data points are in range of tens of thousands and higher in any ML model.

For the quality, the required level of quality for an effective ML model is yet hard to define. One of the reasons is that data has to be unified under certain categories, before the use and only then the quality can be assessed. Before such processing, it is difficult to determine the quality of the data, and often even not possible to grant the quality. This fine-tuning process is called data engineering. Nonetheless, in the end, the required level of quality also depends on the research problem and analysis require.

Regardless of the data quality and quantity, it is important to set a proper goal and research questions for the ML application in order to make the most efficient and effective use. In the following sub-chapter, selected ML use cases will be presented, explaining the interaction between the methods, data and research goals.

Appendix C.5 Machine Learning Use Cases

This section features six use cases of ML, based on published research papers and reports publicly available. Details of each use case are given, each described below:

- **Topic:** the topic of the use case research.
- **Category:** the machine learning type i.e. unsupervised, supervised, reinforced learning and semi-supervised.
- **Methodology:** the machine learning method(s) used in the use case research.
- **Input data:** type of data used that is input for the model.
- **Output:** result data that comes out of the model.
- **Application in the NVA:** how the machine learning method can be used in the context of the NVA.

The list of use cases which are going to be discussed are following:

- Case 1 – Comparison of factors influencing the flight efficiency
- Case 2 – Prediction of pilot misperception on runway excursion risk
- Case 3 – Understanding accident characteristics and contributing factors
- Case 4 – Text classification of occurrence reports
- Case 5 – Public acceptance analysis using classification
- Case 6 – Detection of anomalies in flight operations

Appendix C.5.1 Case 1: Comparison of influencing factors for the flight efficiency

The first use case research [21] tries to identify the greatest factors that contribute to flight inefficiency in the European airspace. The technique proposed uses various machine learning algorithms to combine the biggest number of recorded variables, and determine which factor(s) contributed the most to the flight inefficiency. The main flight efficiency key performance indicators are the average horizontal en-route flight efficiency of the last filed flight plan trajectory and of the actual trajectory. However, it is possible that air navigation service providers are not capable to influence these chosen indicators. Additionally, there are other influencing factors to filed flight plan trajectories, including, for example the coordination with other air navigation service providers and military authorities, airspace

user preferences and external events such as strikes or airspace closure. The research defines the dependency between factors and conducts analysis that is used in order to redesign the airspace structure. The method used in this paper, essentially tries to combine a large dataset containing numerous variables, and find a correlation with a flight (in)efficiency indicator. In a similar matter, defining representative indicators that can help States establish safety targets requires analyses. For example, this method might also be used to identify the most contributing factors to specific safety issues.

Use Case 1 Summary

- **Topic:** Flight Efficiency.
- **Category:** Supervised Learning.
- **Methodology:** Random Forest Regression.
- **Input:** Historical flight data [22], specifically radar tracks.
- **Output:** Ranking of the relative importance of influence factors for the interface HFE (Horizontal en-route Flight Efficiency). With this output, the route structure has the most important influence on the HFE, followed by the direction of the flight. Congestion and daily variability show a low influence.
- **NVA application:** The results of this paper cannot be directly applied to the NVA. However, the use of a regression algorithm to find the most contributing factors to some actual outcome (in this paper this outcome was flight efficiency) may also be used to find the most contributing factor to specific safety issues.

Appendix C.5.2 Case 2: Prediction of pilot misperception on runway excursion risk

This specific study [7] tries to predict which factors are most important for pilots to have a misperception of runway excursion risk. These misperceptions can potentially contribute to a runway excursion since the flight crew fails to recognize and subsequently mitigate the risk(s). The research made use of deidentified flight data publicly available at NASA [37]. The data was used to train a ML-model that predicts whether an approach is exposed to pilot misperception on runway excursion risk based on flight parameters at 500 feet above ground level. In this case, glideslope deviations seem to occur most frequently together with runway excursion risk misperceptions (in other words: continued unstable approaches). The resulting list of factors, which is strongly related to human factor issues, can be used for additional event detection in Flight Data Monitoring (FDM) programs and can be linked to Safety Performance Indicators to manage runway excursion risk.

Use Case2 Summary

- **Topic:** Runway Excursion Risk Perception.
- **Category:** Supervised Learning.
- **Input:** Flight data recorder data collected over a 4 year period of a fleet of 35 regional jets operating in the US. The data were archived and made available by NASA for public use.
- **Output:** Ranking of the misperception of six unstable approach risk factors, which are glideslope deviation, selected approach speed deviation, localizer deviation, flaps not extended, drift angle, and approach speed deviation and ranking of the misperceptions per relevancy.
- **Methodology:** Classification of unstable approaches and ranking of factors that contribute the most to unstable approach risk misperception.

- **NVA application:** Operators may apply this technique to their available flight data to derive a list of factors that have impact on particular accident types, such as runway excursions. The NVA team may use the analysis results as input for the hazard identification and risk assessment process of the NVA.

Appendix C.5.3 Case 3: Understanding accident characteristics and contributing factors

The following research study [23] focussed on the identification of contributing factors in railroad accidents. At a glance, this research might seem very similar to the first use case as it identifies contributing factors to a specific outcome (accidents). However, this specific research makes use of unsupervised text mining, making it essentially different than the first use case since that use case did not use text as input. The data that is used to identify these factors comes from accident and incident reports. Accident and incident reports are provided in the form of documents with text. It is often difficult to extract information from a large number of reports, since it is time consuming to read these reports and manually extract the required information from the different sections of the reports. As a result, it is also time consuming to determine contributing factors to a number of incidents and accidents. Moreover, it is difficult or nearly impossible to manually deduct patterns in accident characteristics and factors across a large number of reports. Text mining and topic modelling can be used in order to automatically discover railroad accident characteristics and contributing factors in those accidents. On the basis of the results, it is possible to predict the costs of extreme accidents and the correlation with injuries or loss of life due to the accidents according to the study [23].

Use Case 3 Summary

- **Topic:** Identifying and understanding accident characteristics and contributing factors.
- **Category:** Unsupervised.
- **Input:** Data available on railroad accidents for the period of 11 years, between 2001 and 2012, which are yearly reports of accidents containing numbers (e.g. accident speed), categorical data (e.g. equipment type) and free text.
- **Output:** the selection of unique words in the 10 topics of accident reports, frequency of 10 topics in the reports, categorization of accident causes.
- **Methodology:** Topic model.
- **NVA application:** Service providers and the authority (ILT/ABL) may apply this technique to their available set of occurrences to detect factors that have played a role in a particular occurrence type. The NVA team may use the analysis results as input for the hazard identification and risk assessment process of the NVA.

Appendix C.5.4 Case 4: Text classification of occurrence reports

This use case proposes the use of a text classifier on an occurrence reporting database in order to automatically classify reported occurrences with a specific taxonomy. By using a supervised text classifier, the classification of thousands of reports can be done in a matter of minutes instead of hours or days when the classification would have to be done by hand. The text classifier uses a supervised learning technique, requiring a training data set. The exact amount of data required for training depends on the number of classes or labels that the classifier must be able to predict, as the training data must have enough training samples per class. The benefit of using such a classifier is two-fold. Firstly, a text classifier has the ability to automatically predict a certain class or label for a particular occurrence

report. Occurrence reporting schemes tend to request a lot of information from the observer, especially reports submitted under the EC 376/2014 regulation. Reporting organizations often struggle to collect and report the required information, therefore some essential information may be underreported. An automatic classifier might be able to automatically assign these missing labels to data fields based on the information that was actually provided in the report. In the use case here, a text classifier was trained to predict the occurrence category of occurrence reports based on the narrative and headline of the report. This approach improves the efficiency of the data processing and contributes to standardization of the labelling of reports. The second benefit of using a text classifier is the ability to perform deeper analysis on a batch of occurrence reports. By for example training the classifier such that it would recognize reports related to a specific safety issue, the data querying for relevant reports related to a specific safety topic can be enhanced. At this moment, querying for these reports is mostly being done based on key-word searches provided by a subject matter expert. However, a classifier can be trained on more elaborate criteria such as specific word combinations and relative word importance to query relevant reports.

Use Case 4 Summary

- **Topic:** Text Classification of Occurrence Reports.
- **Category:** Supervised Learning.
- **Methodology:** Text-mining – Classification.
- **Input Data:** Various incident and accident reports reported to parties such as Civil Aviation Authorities.
- **Output:** Occurrence categories per incident report based on historical data.
- **Application for NVA:** Automatic classification of reports enables more efficient and uniform classification than that of manual classification by (a group of) experts. Analysis of large amounts of reports is done more efficiently, as long as sufficient training data is available to train the model. In this use case, incident reports are labelled with the CICTT Occurrence Category. However, other taxonomies can be used as well. The NVA team may use the analysis results as input for the hazard identification and risk assessment process of the NVA.

Appendix C.5.5 Case 5: Public opinion analysis using classification

Public opinion or (dis)agreement and (lack of) acceptance is of importance for aviation policy development. One way to examine public opinion is to conduct a survey, but this method takes time and resources. One of the ML techniques that can be applied to analyze public opinion is a sentiment analysis. With sentiment analysis it is possible to systematically identify, extract, quantify, and study affective states and subjective information, such as reactions of the general public or stakeholders on specific topics. The NVA could also use public data of, for example, aviation forums or Twitter posts in combination with sentiment analysis to evaluate the public sentiment regarding rules and regulations, safety topics, aviation policy etc. Another example is the application of sentiment analysis on social media information from airport users regarding security procedures or epidemic control measures.

The sentiment analysis technique uses a combination of natural language processing, text analysis, computational linguistics, and biometrics. In a particular study [24] the authors conducted research on sentiments for six US airlines using twitter posts (related to US Airways, United Airlines, American Airlines, Southwest Airlines, Delta Airlines and Virgin America Airlines). The study result is a categorization as well as a visualization of counted/categorized customer complaints of airlines and the polarity of customers on the six airlines in question.

Use Case 5 Summary

- **Topic:** Public opinion analysis.
- **Category:** Supervised learning.
- **Input:** Tweets about six airlines in the United States.
- **Output:** Categorization of sentiment of the writer of the Twitter posts in the categories of: negative, neutral and positive.
- **Methodology:** Sentiment classification.
- **NVA application:** The sentiment analysis on social media related to topics relevant to the NVA may provide the analyst information about the public opinion related to those topics. The topics can be diverse, for example policy, regulations, (future) developments, operational issues, safety, noise complaints. The outcome of the sentiment analysis may be used as input for policy development, promotion campaigns, communications to the public, information/education to the general public and aviation stakeholders.

Appendix C.5.6 Case 6: Detection of anomalies in flight operations

This use case applied two different unsupervised techniques to identify abnormal flight operations of a specific aircraft type [25]. The study made use of flight data from aircraft available in the Flight Data Monitoring program of the operator. The two techniques classify a data sample, i.e. a flight state, as either normal or abnormal. However, the input and output of the two techniques differ slightly. The first method analyzes a specific flight segment of a flight phase. It compares all flights in the data set by specifically analyzing the exact same flight segment. One major drawback of this method is that each segment that will be analysed must have the same number of data points (i.e. segment duration must be equally long). The second method does not have this shortcoming and focusses on analyzing an entire flight for a specific abnormal data point. The study concluded that both methods work and are complementary since they both provide a comparable classification but on a different level of detail. A limitation of the research was the limited amount of data. A large dataset may result in a more robust application of the researched methods.

Use Case 6 Summary

- **Topic:** Detection of anomalies in flight operations based on FDM data.
- **Category:** Unsupervised.
- **Input:** Flight data from one aircraft type. For the first method additional filtering on one specific flight phase was necessary. The phase of the operation as well as length of the frame, i.e. take-off, landing, or 20 seconds after the take-off of all flights must be the same. The second method analysed the entire flight, so no additional requirements were imposed in terms of flight phase or segment length.
- **Output:** Classification of the flight segment as normal or abnormal. The first method only classified whether the entire segment was considered abnormal or not. The second method indicated which specific data point of the entire flight was considered abnormal.
- **Methodology:** Principal Component Analysis.
- **NVA application:** Large scale analysis of flight data to detect abnormal patterns or abnormal (non-nominal) aircraft states in normal flight operations. The method can be applied on FDM data but also on radar track data. The service providers could apply these techniques to their data sets (for example, FDM or radar data sets) to detect potential non-nominal flights or aircraft states. The NVA team may use the analysis results as input for the hazard identification and risk assessment process of the NVA.

Appendix D Zoektermen Twitter API

lelystad airport
 luchtvaart
 luchtvaartveiligheid
 vliegveiligheid
 veiligheid luchtvaart
 overlast luchtvaart
 overlast vliegtuig
 overlast schiphol
 stikstof luchtvaart
 stikstof schiphol
 remkes lelystad
 remkes schiphol
 geluid luchtvaart
 geluid schiphol
 uitstoot luchtvaart
 uitstoot schiphol
 uitstoot lelystad
 uitstoot vliegtuig
 stikstof lelystad
 ovv luchtvaart
 ovv schiphol
 drone regels
 drones
 drone regelgeving
 drone veiligheid
 luchthaven

Airport
 LvnI
 Aanvliegroute
 Vliegroute
 Luchtverkeersleiding
 vlieg veiligheid
 Dronevluchten
 Luchtruim
 klimaatneutraal vliegen
 duurzaamheid luchtvaart
 elektrisch vliegen
 schiphol economie
 Baanonderhoud
 omgevingsraad schiphol
 waterstof vliegen
 hybride vliegen
 duurzame luchtvaart
 schonere vliegtuigen
 emissieloze luchtvaart
 emissieloze luchthavens
 waterstof luchtvaart
 Pipistrel



Dedicated to innovation in aerospace

NLR - Koninklijk Nederlands Lucht- en Ruimtevaartcentrum

Koninklijke NLR werkt als neutraal opererend onderzoekscentrum met zijn partners aan een betere wereld van morgen. NLR biedt daarbij innovatieve oplossingen en technische expertise en zorgt voor een sterke concurrentiepositie van het bedrijfsleven.

NLR is ruim 100 jaar een kennisorganisatie met de diepgewortelde wil om te blijven vernieuwen en zet zich in voor een duurzame, veilige, efficiënte en effectieve lucht- en ruimtevaart.

De combinatie van diepgaand inzicht in de klantbehoefte, multidisciplinaire expertise en toonaangevende onderzoeksfaciliteiten, maakt snel innoveren mogelijk. NLR vormt in binnen- en buitenland de spilfunctie tussen wetenschap, bedrijfsleven en overheid, en overbrugt de kloof tussen fundamenteel onderzoek en toepassingen in de praktijk. Daarnaast werkt NLR als Groot Technologisch Instituut (GTI) sinds 2010 in de TO2-federatie samen aan toegepast onderzoek in Nederland.

Vanuit de hoofdvestigingen in Amsterdam en Marknesse en twee satellietvestigingen, draagt NLR bij aan een veilige en duurzame maatschappij en werkt met partners in vele (defensie)programma's, onder andere aan complexe composieten constructies voor verkeersvliegtuigen en aan doelgericht gebruik van het F-35-jachtvliegtuig. Daarnaast geeft NLR invulling aan Nederlandse en Europese (klimaat)doelstellingen conform de Luchtvaartnota, de European Green Deal, Flightpath 2050 en door deelname aan programma's zoals Clean Sky en SESAR.

Voor meer informatie bezoek: www.nlr.nl

Postal address

PO Box 90502
1006 BM Amsterdam, The Netherlands
e) info@nlr.nl i) www.nlr.org

NLR Amsterdam

Anthony Fokkerweg 2
1059 CM Amsterdam, The Netherlands
p) +31 88 511 3113

NLR Marknesse

Voorsterweg 31
8316 PR Marknesse, The Netherlands
p) +31 88 511 4444