Brugerrejsen - Sådan kommer du i gang

Analyseplatform

Exported on 12/02/2024

Table of Contents

Denne side gennemgår brugerrejsen af et projekt/self service projekt på Analyseplatformen fra projektoprettelsen over adgangsstyringen af projekter og adgang til data, til produktionslægning og dekommisionering.

---

1 Projektoprettelse

For at få adgang til Analyseplatformen skal man oprette et projekt, enten et Self Service[1] projekt eller et internt projekt (Center for Avanceret Analyse, UFST).

Self Service konceptet henvender sig til brugere i de andre styrelser i skatteforvaltningen, samt andre afdelinger i UFST end Center for Avanceret Analyse.

For at kunne udvikle på Analyseplatformen, skal du først have oprettet et Self Service Projekt. Læs mere om self service her[2].

- Oprettelsen af et Self service projekt kan bestilles med denne MitIT-[3]blanket[4].
- Blandt andet skal der angives:
    - **Ansvarlig procesejer (personaleleder)**
    - **Formålet** med at behandle data
    - **Behandlingshjemmel**
    - Hvilke **kategorier af personoplysninger** der indgår i modellen
    - Hvilke **kategorier registrerede** tilhører
    - **Varigheden af behandlingen** samt de forventede tidsfrister for sletning af personoplysninger
    - **Sekundær dobbeltgodkender til Omada**
    - **Self service-niveau**
        - Niveau 1 - ingen brug af Analyseplatform udviklingsserver eller ML Center of Excellence (CoE) infrastruktur
        - Niveau 2 - Adgang til Analyseplatformens udviklingsserver uden produktionslægning
        - Niveau 3 - Fuld adgang til Analyseplatformens infrastruktur med produktionslægning
        - Internt projekt i Center for Avanceret analyse
    - **Oprettelse af adgangsgivende roller (Kun Niveau 2 og 3 samt interne projekter)**
        - For Niveau 2, 3 og interne projekter skal der oprettes en Udvikler-rolle
        - For Niveau 3 og interne projekter skal der desuden oprettes en Bruger-rolle
        - Der er muligt at oprette andre roller hvis der er behov for det.
    - **Oplysninger om projektets første repository**

For niveau 3 projekter vil der efter indmeldelse af projektet blive indkaldt til et møde med data scientists fra UFSTs Machine Learning Center of Excellence (ML CoE). Ved dette møde vil der blive givet hjælp med at

1 https://confluence.ccta.dk/x/Ve2OFg
2 https://confluence.ccta.dk/pages/viewpage.action?pageId=325799287
3 https://skat-myit.onbmc.com/dwp/app/#/srm/profile/AGGJ116OGXG5OARK2QM6R93ATAFOHQ/srm
4 https://skat-myit.onbmc.com/dwp/rest/share/
OJSXG33VOJRWKVDZOBST2U2SIQTHIZLOMFXHISLEHVAUOR2KLBLDKR2JJEYEORCBKI4EIURYKZJDQRCSHBLFCSCYKETHEZLTN52X
EY3FJFSD2QKHI5FDCMJWJ5DVQRZVJ5AVESZSKFGTMURZGNAVIQKGJ5EFCJTDN5XHIZLYORKHS4DFHVBUCVCBJRHUOX2IJ5GUKJT
QOJXXM2LEMVZFG33VOJRWKTTBNVST243SNU======

komme i gang med at bruge platformen, og der vil være mulighed for at udveksle erfaringer med lignende løsninger på platformen.

---

2 Adgangsstyring af projekter på Analyseplatformen

På Analyseplatformen gør vi brug af en rollebaseret adgangsstyring.

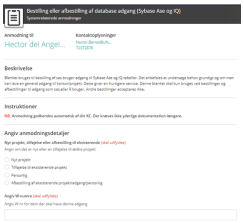
Hvert projekt bliver født med en udvikler- og en bruger-rolle som giver forskellige adgange til projektet og som kan søges igennem Omada[5].

Det er selvfølgelig muligt at tilføje flere roller hvis der er behov for det.

Du kan læse mere om den rollebaserede adgangsstyring på Analyseplatformen her[6].

---

5 http://iam.ccta.dk
6 https://confluence.ccta.dk/x/BBATH



3 Bestilling af adgange til data

Før et projekt kan tilgå data fra data warehouse

- Bestillinger af eksisterende **tabeller i DW** (BI & DW) (med denne MitIT blanket[7])
- **Procesejers eller KCs** ansvar at **godkende** dataadgange til en medarbejder
- Under **udvikling** er det muligt for projekterne at **kopiere data i projektfoldere**
- **Projekterne** er selv ansvarlige for at **rydde op** for data
- **Under træning og produktion forbliver data i DW** – det loades under hvert run

7 https://skat-myit.onbmc.com/dwp/app/#/srm/profile/SRGAA5V0F53B1APF91GKPEMTS7QAOR;providerSourceName=srm

4 Andre bestillinger i AP

- Oprettelse af forskellige pakker baseret på **skabeloner:**
    - **Modelpakker** -> Modeller
    - **Taskpakker** -> Analytics REST API'er
    - **Batchpakker** -> Konfigurerebarer Jenkinspipelines
    - **Analytics Apps** -> Shiny og Streamlit*
- Alle skabeloner indeholder en **metadatafil** som indeholder oplysninger om projektet. Dette bruges til at danne AP's **applikationsoverblik**
- Bestilling af **support**

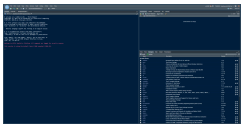
Bestillinger deployment samt supporthenvendelser laves igennem MitIT med denne blanket[8].

Bestillinger vedrørende projekter, f.eks. oprettelse af nye repositories eller ændringer i roller laves igennem denne blanket.[9]

---

8 https://skat-myit.onbmc.com/dwp/rest/share/OJSXG33VOJRWKVDZOBST2U2SIQTHIZLOMFXHISLEHVAUOR2KLBLDKR2JJEYEORCBKI4EIURYKZJDQRCSHBLFCSCYKETHEZLTN52XEY3FJFSD2QKHI5BEONRYKZBUUQSIGFAVEOKLKVIVSURYJREDGQSKJE4FKJTDN5XHIZLYORKHS4DFHVBUCVCBJRHUOX2IJ5GUKJTQOJXXM2LEMVZFG33VOJRWKTTBNVST243SNU======

9 https://skat-myit.onbmc.com/dwp/app/#/srm/profile/AGGJ116OGXG5OARK2QM6R93ATAFOHQ/srm



5 Udvikling

- Udvikling i **R** på en **serverbaseret IDE** - Rstudio
- Udvikling i **python** både på en **serverbaseret IDE** eller via en lokal **Vscode med ssh** forbindelse til AP
- **AP 3 CI/CD** bygge- og deployjobs samt træning som en del af byggejobbene
- **AP 4 CI/CD** bygge- og deployjobs – Afkoblet træning i **MLFlow**
- **CI/CD** for at sikre blandt andet kvalitet og reproducerbarhed
- **Continuos delivery** sikrer i dag **funktionsadskillelse** og er processen for **change management** – på sigt kan vi opnå **continuos deployment**

5.1 Compliancekrav

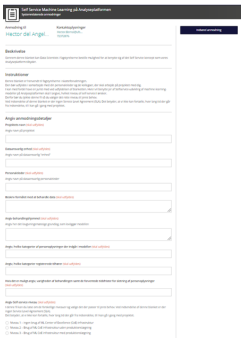
Compliancekrav som ikke er implementeret eller kan implementeres på AP skal afdækkes af projekterne i denne skabelon[10].

- Udfyldt compliance blanketter:

1. Stamoplysninger
2. Databeskyttelse by design
3. Forvaltningsret by design
4. Informationssikkerhed
5. Øvrige lovkrav

---

10 https://confluence.ccta.dk/display/MLkrav/Lovmedholdelighed+og+compliance



6 Produktionslægning

- **Intern MDI**
  - Review intern i MDI-teams
  - Deployments (change requests) bestilles via MitIT (med denne blanket[11])
- **Self Service**
  - AA reviewer ML-projekter udviklet på analyseplatformen, hvis de indeholder en ML- eller AI-model. OBS, hvis en shiny-model udstiller data fra datawarehouse, men ikke foretager en AI eller ML-behandling af data må vi ikke produktionslægge den.
- **Compliancekrav** skal være på plads inden produktionslægning og verificeres som en del af review

6.1 Drift

Før produktionslægning af en ny ML-løsning bør et årshjul, som eksemplet i dokumentet Årshjul for ML-løsninger i produktion.pptx[12], udarbejdes for den løbende kvalitetssikring og vedligehold af løsningen. Udarbejdelsen af årshjulet inkluderer fastlæggelse af de løsningsspecifikke rammer for hver af de listede driftsaktiviteter i tabellen til højre (ikke alle aktiviteter er obligatoriske).

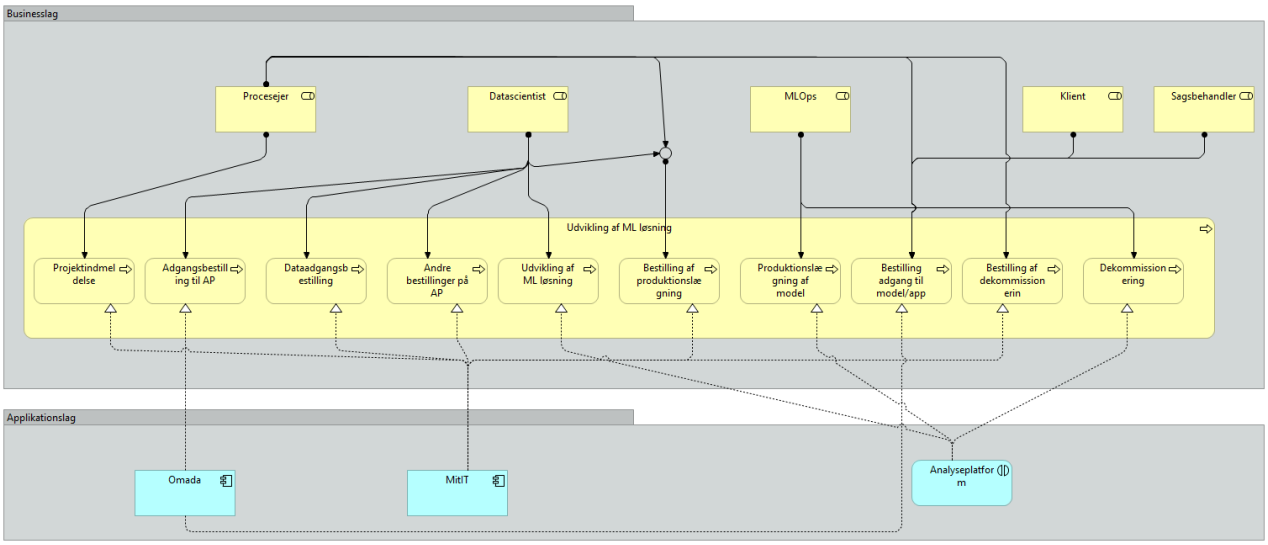
11 https://skat-myit.onbmc.com/dwp/app/#/srm/profile/AGGBG68VCJBH1AR9KUQYR8LH3BJI8U;providerSourceName=srm

12 https://confluence.ccta.dk/download/attachments/475442535/%C3%85rshjul%20for%20ML-l%C3%B8sninger%20i%20produktion.pptx?api=v2&modificationDate=1702024869309&version=1

7 Dekommissionering

- Projekterne bestiller dekommissionering via MitIT[13]
- De forskellige artefakter som udgør en applikation slettes eller arkiveres

Brugerrejsens processer og involverede applikationer



13 https://confluence.ccta.dk/x/HC1wEw

8 Adgangsstyring af projekter på Analyseplatformen

8.1 Rolle baseret adgangsstyring af projekter på Analyseplatformen

Den rolle baserede adgangsstyring på analyseplatformen er bygget rund omkring de enkelte projekter frem for analyseplatformen som helhed.

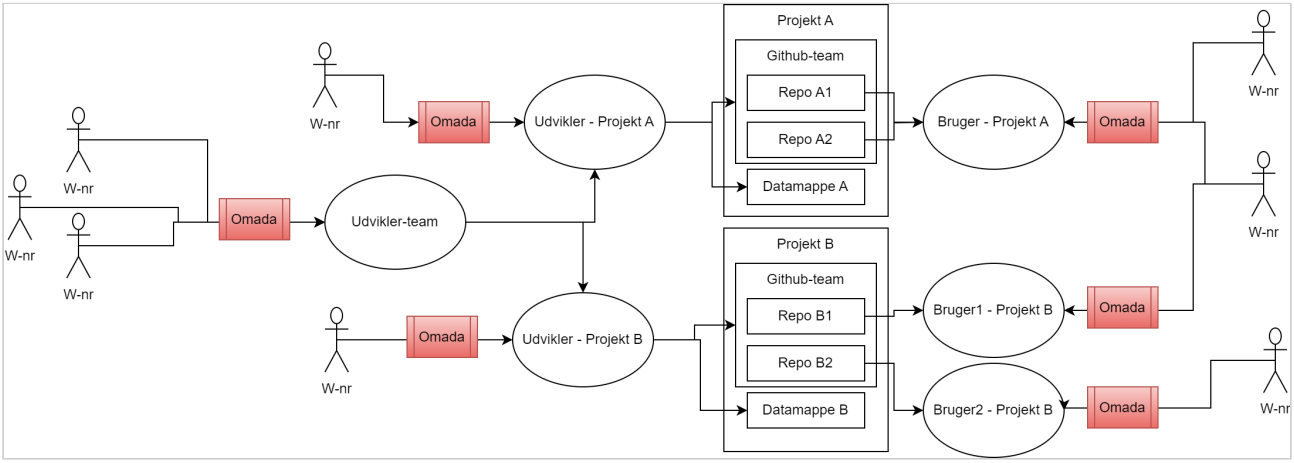
Som bruger af Analyseplatformen søger du ikke adgang til Analyseplatformen selv, men en adgang til udvikler-rollen af et projekt der bliver udviklet på platformen.

Adgangen til udvikler-rollerne, lige som bruger-rollerne, for de enkelte projekter søges i Omada,[14]hvor de ligger under systemet "Analyseplatform APPs".

Adgangen til Udvikler-rollen vil give en udvikler alle de adgange der er nødvendigt for at udvikle på et given projekt, herunder (afhængig af projektets self service niveau) adgang til udviklingsserveren, projektets datamappe, github-team, repositories, jenkins-pipelines og test-miljøer.

Ud over de projektspecifikke udvikler-roller er det muligt at oprette **udvikler-teams.** Udvikler teams er tiltænkt kontorer/teams der arbejder på tværs af flere forskellige projekter.

Hvis du ønsker at tildele læserettighed til jeres kode for et konkret repositorie og en konkret kollega skal dette anmodes gennem supportblanketten du finder på denne side[15]



14 https://iam.ccta.dk/dashboard.aspx
15 https://confluence.ccta.dk/x/He2OFg

9 Bestillinger, fejlmeldinger og support på Analyseplatformen

Analyseplatformen benytter MitIT til følgende opgaver:

Herunder får du en vejledning til, hvordan du anmoder om de forskellige former for service.

Inden du fejlmelder er det dog vigtigt, at du har afsøgt følgende muligheder:

- Google, StackOverflow og lignende
- Support hos Data Scientist-kolleger
- Dokumentation for Analyseplatformen

Ydermere kan du også søge assistance blandt Data Science kollegaer i dit team, eller i data science sparring kanalen[16] på teams.

Vær opmærksom på at Analyseplatformen ikke er ansvarlig for adgang til data warehouse og data i data warehouse. Adgang til data warehouse kan ansøges gennem denne blanket[17]. Analyseplatformens udviklingsmiljø trækker data fra KDW Preprod, data I KDW Preprod bliver kopieret ind fra KDW Prod hver fredag fra klokken 22 og er typisk færdig engang lørdag morgen. Hvis jeres projekt kræver data opdateret med hyppigere interval (fx dagligt) skal I bestille det også gennem blanketten her i paragraffen. Hvis du har brug for adgang til prod med det formål at debugge en produktionslagt model bestilles dette gennem supportblanketten.

9.1 **Fejlmelding**

Du fejlmelder fejl du oplever på Analyseplatformen gennem denne blanket[18]

Når du får siden op vælger du "Andre programmer/systemer", derefter vælger du "Analyseplatform" i dropdown. Derefter udfylder du de sidste felter.

16 https://teams.microsoft.com/l/channel/19%3ahGtyyeN2sJ_bPWFgw1c_Jq_e5eUByHalisTwuItxhDk1%40thread.tacv2/Generel?groupId=c87b5349-e2de-403a-839a-a328f4e4dc2e&tenantId=2e93f0ed-ff36-46d4-9ce6-e0d902050cf5

17 https://skat-myit.onbmc.com/dwp/rest/share/OJSXG33VOJRWKVDZOBST2U2SIQTHIZLOMFXHISLEHVAUOR2KLBLDKR2JJEYEORCBKI4EIURYKZJDQRCSHBLFCSCYKETHEZLTN52XEY3FJFSD2U2SI5AUCNKWGBDDKM2CGFAVARRZGFDUWUCFJVKFGN2RIFHVEJTDN5XHIZLYORKHS4DFHVBUCVCBJRHUOX2IJ5GUKJTQOJXXM2LEMVZFG33VOJRWKTTBNVST243SNU======

18 https://skat-myit.onbmc.com/dwp/rest/share/OJSXG33VOJRWKVDZOBST2U2SIQTHIZLOMFXHISLEHUYDAMBQGAYDAMBQGAYDAMJQGMTHEZLTN52XEY3FJFSD2QKHI5BEOMJZGBBE4SRZINAVCNRSGBFVSUJVGRATQUCXKI4UIJTDN5XHIZLYORKHS4DFHVBUCVCBJRHUOX2IJ5GUKJTQOJXXM2LEMVZFG33VOJRWKTTBNVST243SNU======

9.2 Fejlmelding af Machine Learning Løsning (vi er ved at finde ud af den rigtige måde at fejlmelde machine learning løsninger på)

Denne blanket er ved at blive opdateret. Formålet med blanketten er, at brugere af modellerne (sagsbehandlere i styrelserne) kan fejlmelde en machine learning løsning, hvis den ikke virker efter hensigten. Det kan ske her[19]

Blanketten må godt benyttes til formålet i dens nuværende form. Hvis man ønsker at fejlmelde en model, der ikke er på listen kan man anvende "andet" i dropdown.

9.3 Support eller deployment af model

Søges gennem denne blanket[20]

9.4 Oprettelse af Projekter/Selfservice projekter på Analyseplatformen, tilføjelse af repositories eller roller til eksisterende projekt, tildeling af adgang

Gennem denne blanket[21] kan du søge de forskellige projektrelaterede services, som muliggør, at du kan arbejde på Analyseplatformen. Ved spørgsmål ang. blanketten må du godt kontakte gennem Analyseplatformen Support-kanal[22], ved spørgsmål ang. anvendelsen af Analyseplatformen indsend da supportblanketten ovenfor.

Du kan læse om Self-service her Self-service på Analyseplatformen[23]

9.4.1 Særligt om Shiny-apps

Vær opmærksom på at brugen af Shiny-app kun kan benyttes til nedenstående use-cases:

1. Når der udvikles løsninger, som skal understøtte real time ML modelscoringer og scoringer på enkeltpersoner.

---

19 https://skat-myit.onbmc.com/dwp/rest/share/OJSXG33VOJRWKVDZOBST2U2SIQTHIZLOMFXHISLEHUYDAMBQGAYDAMBQGAYDAMJQGMTHEZLTN52XEY3FJFSD2QKHI5AUUNKCGREUQVCTIJAVCWKDKVFFOUKYIRMTSRRVJ5MEWJTDN5XHIZLYORKHS4DFHVBUCVCBJRHUOX2IJ5GUKJTQOJXXM2LEMVZFG33VOJRWKTTBNVST243SNU======

20 https://skat-myit.onbmc.com/dwp/rest/share/OJSXG33VOJRWKVDZOBST2U2SIQTHIZLOMFXHISLEHUYDAMBQGAYDAMBQGAYDAMJQGMTHEZLTN52XEY3FJFSD2QKHI5BEONRYKZBUUQSIGFAVEOKLKVIVSURYJREDGQSKJE4FKJTDN5XHIZLYORKHS4DFHVBUCVCBJRHUOX2IJ5GUKJTQOJXXM2LEMVZFG33VOJRWKTTBNVST243SNU======

21 https://skat-myit.onbmc.com/dwp/rest/share/OJSXG33VOJRWKVDZOBST2U2SIQTHIZLOMFXHISLEHUYDAMBQGAYDAMBQGAYDAMJQGMTHEZLTN52XEY3FJFSD2QKHI5FDCMJWJ5DVQRZVJ5AVESZSKFGTMURZGNAVIQKGJ5EFCJTDN5XHIZLYORKHS4DFHVBUCVCBJRHUOX2IJ5GUKJTQOJXXM2LEMVZFG33VOJRWKTTBNVST243SNU======

22 https://teams.microsoft.com/l/channel/19%3a3bf36fa6bc1a46b2b685b744b8b3f172%40thread.tacv2/Analyseplatform%2520Support?groupId=7ea7a7d4-f0da-4647-8eb4-6c1fec726d69&tenantId=2e93f0ed-ff36-46d4-9ce6-e0d902050cf5

23 https://confluence.ccta.dk/pages/viewpage.action?pageId=325799287

2. Når der er behov for at udstille forklaringer til ML modelscoringerne og til ML modellens performance.
3. Når der er behov for interaktive dashboards, der gør brug af avanceret statistisk funktionalitet.
4. Shiny apps kan benyttes i dag til indsamling af data til machine learning modeller. Det vurderes dog fra case til case, om denne indsamling kan foregår, evt. hvor data gemmes i en database frem for via excel-ark, eller om denne indsamling kun skal foregå via den nyudviklede Qlik Sense app til feedback.

Shiny apps må ikke benyttes til almindelig BI rapportering. Det behov dækkes af BO og Qlik Sense.

10 Idriftsættelse af modeller på Analyseplatformen

10.1 Informationssikkerhed - vejledning til projekter

Denne side indeholder vejledning til at udfylde skemaet fra Informationssikkerhed (ISK)

Retningslinje for sikkerhed ved brug af data og analyse (Informationssikkerhedsportalen)[24]

Lovkrav og compliance for machine learning[25] indeholder en samling skemaer klar til udfyldelse for det enkelte projekt.

10.1.1 Vejledning

1. Lav en kopi af confluence-siden[26] og placer denne på en valgt placering under projektet.
2. Udfyld kopien - nedenstående vejledningstekster kan være til hjælp

10.1.1.1 Vejledningstekster

I det nedestående skema beskriver vi i hvilket omfang korrekt brug af analyseplatformen medfører at vi mener projektet opfylder kravet. Projektet må selv beskrive yderligere tiltag eller projektspecifikke grunde til afvigelser fra kravet.

I fortolkning af kravene kan følgende slides fra sikkerdigital være til yderligere hjælp

sikkerdigital.dk - tiltag til at sikre brugen af kunstig intelligens[27]

Disclaimer

Denne side indeholder ikke nødvendigvis den aktuelle liste og beskrivelse af krav. Opsøg altid kilden fra Informationssikkerhed for at få den seneste version.

ID	Krav	Løst af plat-formen?	Vejledningstekst	Udfyldes senest

24 https://skat.sharepoint.com/sites/Informationssikkerhed/SitePages/Politikker,%20standarder%20og%20retningslinjer.aspx
25 https://confluence.ccta.dk/display/MLkrav/4.+Informationssikkerhed
26 https://confluence.ccta.dk/display/MLkrav/4.+Informationssikkerhed
27 https://sikkerdigital.dk/media/6839/vejledning-tiltag-til-at-sikre-brugen-af-kunstig-intelligens-2020.pdf

	Krav til træning i sikker brug af AI			
2.1	Medarbejdere, der arbejder med AI, skal have tilstrækkelig viden omkring sårbarheder ved og trusler mod brugen af AI-baserede løsninger. Dette kan fx opnås gennem sidemandsoplæring, introforløb eller anden form for uddannelse. Træningen skal være gennemført inden arbejdet påbegyndes.	Ja	Kontoret Avanceret Analyses ansvar: Nye data scientists skal onboardes i: Retningslinje for sikkerhed ved brug af data og analyse (Informationssikkerhedsportalen)[28], Informationssikkerhedshåndbogen (Informationssikkerhedsportalen)[29], ML-materiale fra Persondataportalen[30], AA Best practice[31], Brug af Analyseplatformen (Under udarbejdelse)	Inden udvikling
2.2	Oplæringen af nye medarbejdere skal indeholde fokus på den enkeltes ansvar (fx at rapportere mulige hændelser, sikre best practice datahåndtering, være opmærksom på relevante trusler, udføre kontrol af modeloutput mv.)	Ja	Se 2.1	Inden udvikling
	Krav til backup og gendannelsesplan			

28 https://skat.sharepoint.com/sites/Informationssikkerhed/SitePages/Politikker,%20standarder%20og%20retningslinjer.aspx
29 https://skat.sharepoint.com/sites/Informationssikkerhed/SitePages/Informationssikkerhedsh%C3%A5ndb%C3%B8ger.aspx?web=1
30 https://skat.sharepoint.com/sites/perdata/SitePages/Persondataret%20og%20it-udvikling.aspx
31 https://confluence.ccta.dk/display/daiufst/Best+Practice+for+Data+Scientists+i+Avanceret+Analyse

3.1	Der skal foretages jævnlige backups af kode og data, og disse skal følge retningslinjen for backup.	Ja	Hvis projektet benytter platformen som foreskrevet, er dette krav efterlevet. Platformen understøtter kravet gennem: Github – kode (DevOps kontoret), CLS – logs (DevOps kontoret), UFSTs DW – data (BI&DW DevOps), Serverer – data og platform (Infrastruktur og cloud), Gendannelsplan – AP (ansible + manuel gendannelse + Lifecycle Management)	Inden produktions -lægning
	Krav til datakryptering			

4.1	AI Data skal som udgangspunkt krypteres i overensstemmelse med retningslinjen for kryptering. Dog kan AI modeller ikke trænes på krypteret data, så hvis dette er nødvendigt, skal projektet definere en fast cyklus, hvor data dekrypteres, så træningen kan gennemføres. Dette kan være en gang månedligt, eller med andre frekvenser, alt efter risikoniveauet samt behovet for træning. Alternativt kan homomorfisk kryptering eller hashing anvendes, da begge metoder tillader at træne modellen og forudsige udfald på det behandlede data uden at gå på kompromis med krypteringen.	Delvist	Hvis projektet benytter platformen som foreskrevet, er dette krav efterlevet. Platformen understøtter kravet gennem: Data ligger krypteret på harddiskene i DataWarehouse. Træningen startes af en udvikler eller anden skattemedarbejder men udføres af platformens automatik. Under træning behandler modellen data uden kryptering. Denne læsning af data foretages via en service account som har minimal dataadgang. Træningsdata overbevares ikke i platformen efter endt træning. Sammen med platformens øvrige beskyttelse opfattes denne opsætning som tilstrækkelig sikker. Der mangler kryptering under transit, men DW understøtter ikke dette jf. Informationssikkerhedsportalen - Skatteministeriets krypteringsstandard.pdf[32]	Inden produktions-lægning
	Krav til penetrationstest af AI-systemer			

32 https://skat.sharepoint.com/sites/Informationssikkerhed/SiteAssets/Forms/AllItems.aspx?id=%2Fsites%2FInformationssikkerhed%2FSiteAssets%2FSitePages%2FH%C3%A5ndb%C3%B8ger%2D%2Dpolitikker%2Dog%2Dretningslinjer%2FSkatteministeriets%20krypteringsstandard%2Epdf&parent=%2Fsites%2FInformationssikkerhed%2FSiteAssets%2FSitePages%2FH%C3%A5ndb%C3%B8ger%2D%2Dpolitikker%2Dog%2Dretningslinjer

5.1	Mindst én gang årligt, eller ved væsentlige ændringer, skal der planlægges og udføres uvildige penetrationstest af anvendte AI-systemer, således at sikkerhedsniveauet testes, og sårbarheder identificeres. Projektet skal bestille penetrationstests ved kontoret Operationel IT-sikkerhed	Ja	Platformen penetrationstestes årligt.	-
5.2	Der skal foretages sårbarhedsscanning af anvendte AI-løsninger, med bistand fra kontoret Operationel IT-Sikkerhed, i tråd med den eksisterende sårbarhedsscanningspolitik .	Ja	Samme som 5.1	-
5.3 (kun relevant ved høj risiko)	Der skal testes flere typer simulerede angreb mod AI-modellen, herunder forsøg på ændret kode, forgiftning af data, data extraction etc. Dette for at opbygge og træne passende reaktionsmønstre overfor kendte typer angreb.	Nej	AP understøtter ikke udvikling af højrisiko modeller.	
	Krav til styring af AI-modeludvikling og -træning			

6.1	AI-modellerne skal versionsstyres så det gøres muligt at spore og dokumentere ændringer i modelkoden, at flere kan arbejde på de samme filer samtidig, og at tidligere versioner af kode kan gendannes i tilfælde af fejl, brud på kodens integritet eller behov for dokumentation.	Ja	Platformen understøtter dette hvis Avanceret Analyses standarder efterleves, herunder brug af Github og model-template.	Inden produktions-lægning
6.2	Der skal inden udvikling af en model-løsning som en del af security-by-design indtænkes valg af eventuel validitet-forøgende tiltag, herunder ”label smoothing”, ”ensemble modeller” og/ eller ”defensiv destillering”. Alle tre begreber er beskrevet i definitionsbilaget.	Nej	Op til projektet, men sjældent relevant for modeller der kun udstilles internt.	Inden produktions-lægning
	Krav til sikring af modellens robusthed			
7.1	Der skal til enhver AI-model udpeges en modelejer med ansvar for hele modeludviklingen, og denne er ansvarlig for at der opstilles kriterier til evaluering af modelperformance, fx angående præcision og fejlrate.	Nej	Op til projektet men påkrævet for at starte et projekt på AP	Inden udvikling
7.2	Det er modelejeren der, sammen med projektet og eventuel styregruppe, fastsætter træning og testforløb for den givne AI-model, og tilsikrer at disse overholdes.	Nej	Op til projektet	Inden produktions-lægning

7.3	Der skal foretages kvalitetssikring af modellens robusthed løbende under hele modeludviklingen og med hyppigt interval, da det øger modellens evne til at håndtere ny, uset data, og dermed gør den sværere at manipulere.	Nej	Op til projekterne men årshjul påminder om dette	Inden produktions-lægning
7.4	Der skal fra starten af AI-modeludvikling indtænkes, at man bevidst arbejder med at adressere så mange sårbarheder som muligt, og derved besværliggøre angreb. Denne iteration af sårbarhedshåndtering bør følge kvalitetssikringen fra krav 7.1.	Delvist	Platformen yder en generel beskyttelse og er risikovurderet. Projektet bør selv afdækker yderligere risici.	Inden produktions-lægning
7.5	AI-Modeller skal testes inden idriftsættelse, for at afklare bias og teste om data kan bære modelhypotesen, og skal godkendes af projektet og eventuel styregruppe, bl.a. på baggrund af de opstillede performancekrav, herunder fx det accepterede antal falske positive resultater, inden modellen flyttes til produktion.	Nej	Op til projektet	Inden produktions-lægning
	Krav til mitigering af læk af modelparametre og -beregninger			

8.1	I forbindelse med design eller tilvejebringelse af en AI-model skal der internt i projektet laves et overblik over, hvor meget modeloutput der vil blive offentliggjort. Der skal i den sammenhæng tilstræbes at nedbringe mængden af modeloutput og viden om modelparametre til et absolut minimum. Dette gør det sværere for hackere at kortlægge den bagvedliggende data og algoritme.	Nej	Modeloutput offentliggøres som hovedregel ikke. Hvis et projekt har behov for at udstille dele af en models output til offentligheden, eks. via skat.dk[33], skal projektet aktivt tage stilling til dette punkt. Udstilling internt i SKAT anses for at være tilstrækkeligt beskyttet under generelle retningslinjer for dataminimering.	Inden produktions -lægning
8.2	Der skal i dialogen omkring kvalitetssikring laves bevidste overvejelser og valg om brug af mitigerende tiltag som fx maskering af output data i aggregerede værdier frem for eksakte værdier, begrænsning i anvendelsesfrekvensen af modellen eller andre tiltag, der kan besværliggøre misbrug af output og reverse engineering af modellen.	Nej	Projektet bør tage stilling til dette, men det vil ofte være nødvendigt at undlade konkrete initiativer da dette vil hæmme sagsbehandlingen (der henvises til skattekontrolloven).	Inden produktions -lægning
8.3	Funktionalitet i modellen, som eksempelvis hyperparametre og lignende information der kan misbruges, må ikke gøres synlige for eksterne.	Nej	Projektet skal efterleve dette krav.	Inden produktions -lægning
	Krav til begrænsning af inputdata			

33 http://skat.dk

9.1	En anvendt AI-model skal under design og idriftsættelse pålægges begrænsninger på tilladt inputdata. Afhængig af modellens natur opsættes begrænsningerne ud fra en logisk betragtning om modellens anvendelsesområder. Dette gøres for, ikke at inkorporere inputmuligheder til modellen i form af umuligt forekommende data, fx en persons anciennitet kan ikke overstige samme persons alder.	Nej	Platformen understøtter brug af denne funktionalitet (via daRecorder og/eller input validation, eks. jsonvalidate). Det påhviler projektet at benytte denne eller redegøre for hvorfor dette er fravalgt. Kan besvare med DPIA (negativ moms)	Inden produktions -lægning
9.2	Modellens typer af inputdata skal kortlægges, og risikoen for ondsindet udnyttelse af dataene, fx hvis der er særlige karakteristika tilknyttet datatypen, vurderes for hvert input. På baggrund heraf skal der foretages en objektiv analyse af mulige restriktioner, herunder bl.a. en analyse af datainputtets mulige værdier i et anvendelsesmæssigt perspektiv, og hvilke af disse restriktioner der skal implementeres.	Nej	Se 9.1	Inden produktions -lægning
	Krav til overvågning af input og output			

10.1	Modelinput og -output skal systematisk monitoreres, for at reducere sandsynligheden for, at ondsindede eksterne aktører misbruger AI-modellen allerede under modeltræning og efter idriftsættelse.	Nej	Planlægges som en del af platformens modelovervågning. Den beskrevne risiko i kapital 10 vurderes kun relevant for projekter der udstiller modellen til offentligheden, ikke til løsninger udstillet internt i SKAT.	Inden produktions-lægning
10.2	Input til og output fra AI-modellen skal screenes for anomalier, hvor projektet har vurderet at det giver mening. Anomalier kan være fx ændringer i frekvensen af specifikke udfald, distributionen af data eller antallet af forespørgsler. Hvis ikke der eksisterer en metode til screening og validering af de pågældende data, skal en sådan oprettes. Der kan være situationer hvor det ikke vurderes til at give mening, hvis fx det rå output ikke vises til brugeren, eller hvis modellen ikke længere lærer fra input.	Nej	Op til projektet at vurdere relevansen af dette krav. Kan typisk fraviges for løsninger internt i SKAT.	Inden produktions-lægning
10.3 (kun relevant ved høj risiko)	Når nyt inputdata introduceres, skal distribution og karakteristika holdes op mod et 'rent' datasæt med kendte egenskaber. Det kunne fx være det nuværende træningsdata eller et andet renset datasæt, som kun bruges til validering af nyt data. Se i øvrigt afsnit 4 omkring krav til kryptering af data i hvile i forhold til beskyttelse af det ”rene datasæt”.	Nej	AP understøtter ikke udvikling af højrisiko modeller.	

10.4 (kun relevant ved høj risiko)	Rene datasæt skal monitoreres og vedligeholdes løbende, sådan at man altid er i kontrol med, hvorvidt datasættet forbliver rent. Det kunne være gennem validering af afvigelser på hash keys, filstørrelse, osv. Skulle der være mistanke om afvigelser, altså forurening af datasæt, skal der iværksættes en udskiftning til et nyt, rent datasæt.	Nej	AP understøtter ikke udvikling af højrisiko modeller.	
10.5 (kun relevant ved høj risiko)	Output skal screenes for tegn på, at modellen er blevet kompromitteret. Det kunne være ved at sammenligne med historisk output. Store forskelle mellem output af benchmarking-og produktionsmodellen kan være tegn på manipulation og bør undersøges nærmere af projektet. Se i øvrigt afsnit 4 omkring krav til kryptering af data i hvile i forhold til beskyttelse af historisk output.	Nej	AP understøtter ikke udvikling af højrisiko modeller.	
10.6	Hvis anomalier identificeres, skal data renses, og modellen gentrænes om nødvendigt.	Nej	Projektet bør tage stilling til dette krav. Det kan ofte fraviges med begrundelsen at kerneopgaven er at opdage anomalier, hvilket kan være tegn på skattesvig (jf. skattekontrolloven).	Inden produktions-lægning
	Krav til benchmarking af model			

11.1	For at benchmarke AI-løsningen skal outputtet af AI-modellen sammenlignes med andre, validerede metoder til at frembringe samme type output. Dette skal som udgangspunkt foregå både under træning og løbende mens modellen er i brug. Det skal dog af ressourcemæssige årsager vurderes, hvorvidt det er nødvendigt at foretage benchmarking løbende, mens modellen er i brug, eller om det er tilstrækkeligt med enkelte stikprøver i brugsfasen. Vurderingen bør bl.a. bero på risikovurderingen af den anvendte platform, kritikaliteten af modellens behandlede data og kritikaliteten af modellens output. Med validerede metoder menes her en tidligere version af AI-modellen, en anden AI-model trænet separat på samme træningssæt af data, eller en simplere, mere transparent model.	Nej	Projektet skal definere en årlig evaluering af data drift og eventuelt afvigelse fra forventning forsøges korrigeret. Se sikkerdigital.dk - tiltag til at sikre brugen af kunstig intelligens[34].	Inden produktions-lægning
11.2	Der skal benchmarkes mod en metode, som bygger på ikke-manipuleret output, og det er derfor vigtigt, at benchmarket ikke selv er kompromitteret. Se afsnit 4 omkring krav til kryptering af data i hvile i forhold til beskyttelse af benchmark output data.	Nej	Op til projektet at vurdere relevansen af dette krav. Kan typisk fraviges for løsninger internt i SKAT. Se sikkerdigital.dk - tiltag til at sikre brugen af kunstig intelligens[35].	Inden produktions-lægning

34 https://sikkerdigital.dk/media/6839/vejledning-tiltag-til-at-sikre-brugen-af-kunstig-intelligens-2020.pdf
35 https://sikkerdigital.dk/media/6839/vejledning-tiltag-til-at-sikre-brugen-af-kunstig-intelligens-2020.pdf

	Krav til træning på manipuleret (adversarial) data			
12.1 (kun relevant ved høj risiko)	Inkludér proaktivt eksempler på ”adversarial data” i det datasæt, AI-modellen trænes på, da det øger modellens robusthed overfor denne type af angreb. Det er vigtigt at disse bevidst indlejrede adversarial data associeres med korrekte labels, AI-modellen derfor lærer at genkende.	Nej	AP understøtter ikke udvikling af højrisiko modeller. Ofte kan man desuden argumentere for at løsningen er delvist beskyttet ved at bruge DataWarehouse som datakilde.	
12.2 (kun relevant ved høj risiko)	Inputdata skal så vidt praktisk muligt modificeres, så der tilføjes små ændringer eller tilfældigheder til data, da det gør det mere besværligt for hackere at manipulere output til egen fordel. Denne øvelse kræver specialistviden at udføre, da den rette balance af ”støj” skal tilføjes til inputdata således den ønskede robusthed opnås uden at ødelægge modellens evne til at tolke korrekt på input data.	Nej	Samme som 12.1	
12.3 (kun relevant ved høj risiko)	Det er vigtigt når støj implementeres, at modellens performance testes. Hvis implementering viser sig at påvirke den generelle modelpræcision og -ydeevne unødvendigt, skal modificeringen revurderes, justeres og gentestes, indtil det acceptable resultat opnås.	Nej	Samme som 12.1	

12.4 (kun relevant ved høj risiko)	Effekten af modificeringen bør løbende evalueres, for at nedbringe risikoen for at efterlade manipuleret data, der med tiden forringer kvaliteten af outputdata.	Nej	Samme som 12.1	
12.5 (kun relevant ved høj risiko)	Der skal i udarbejdelsen af eksemplerne på adversarial data foretages en risikobaseret kortlægning af den inputtype eller –kombination, hvor AI-modellen er særligt sårbar overfor et adversarial attack. Dette gøres af udviklere og evt. eksterne specialister, og resultatet dokumenteres.	Nej	Samme som 12.1	
12.6 (kun relevant ved høj risiko)	Kravet er særligt gældende for AI-løsninger med ”hårde labels”, hvilket vil sige labels er binært tildelte (fx ja/nej, 0/1). Her skal anvendes ”label smoothing”, ”ensemble modeller” og/eller ”defensiv destillering” med to modeller. Det skal vurderes ved hver AI-løsning med hårde labels hvilken kombination af validitet-forøgende tiltag der er bedst egnet.	Nej	Samme som 12.1	
	Krav ved brug af Robotic Process Automation (RPA)			

13.2	Inden idriftsættelse, skal der udføres kvalitetssikring af de datasæt der påtænkes at anvendes RPA på. Der skal evalueres på data klassificering, adgangskontrol, risiko og konsekvens ved læk eller manipulation, mulighed for monitorering osv. PDB skal på baggrund af evalueringen rådgive omkring hvorvidt datasættet må anvendes.	Nej	AP understøtter ikke RPA.	
13.4	Så vidt muligt, skal MFA implementeres sammen med brugernavn og password godkendelse på RPA-operationer.	Nej	AP understøtter ikke RPA.	
13.5	Det skal tilsikres, at login-oplysninger til en anvendt robot skiftes jf. Skatteministeriets retningslinje for Adgangsstyring, og så vidt muligt pålægges at lagres i en Privileged Access Management (PAM) løsning, sådan at overvågning af brug, kendskab til kodeord og skift af kodeord styres automatisk og dermed sænker risikoen for misbrug.	Nej	AP understøtter ikke RPA.	

13.6 (kun relevant ved kritiske systemer)	Kildekoden til anvendte robotter skal scannes for sårbarheder løbende. Der skal så vidt muligt planlægges og foretages penetrationstest med henblik på specifikt at teste kildekodens sårbarhed for målrettede udefrakommende angreb. Penetrationstests bestilles ved OIS.	Nej	AP understøtter ikke RPA.	
13.8	Der skal implementeres validering af in- og outputdata, samt foretages løbende systematisk monitorering af disse, sådan risikoen for misbrug mindskes. Adgang til inputdata skal begrænses til nøje udvalgte, privilegerede brugere. Input skal for øvrige brugere være låst når en robot er aktiv, for at mindske risikoen for uautoriseret manipulation. Outputdata skal valideres for tegn på netop manipulation, fejlberegninger og systemiske fejl. Hvis der findes tegn på disse, skal den gældende proces for sikkerhedshændelser anvendes.	Nej	AP understøtter ikke RPA.	
13.9	Robot kildekode, input data og output data skal krypteres i overensstemmelse med retningslinjen for kryptering. Så vidt praktisk muligt, skal data enten krypteres eller hashes under behandling.	Nej	AP understøtter ikke RPA.	

13.10	Der skal inden idriftsættelse af en robot, fastsættes det nødvendige sæt adgange. Roller tildeles efterfølgende til alene at udføre de nødvendige handlinger. Et eksempel kunne være, hvis en robot skal hente data ét sted, og sende dem et andet sted hen, vil læseadgang på kildelokationen være tilstrækkelig, hvorfor skriveadgang fjernes.	Nej	AP understøtter ikke RPA.	
13.11 (kun relevant ved kritiske systemer)	Så vidt muligt anvendes Privileged Session Management (PSM) på robot opgaver, således adgangsstyring, brug og udført arbejde overvåges. PSM opnås ved at anvende "Vaulting" teknologi til at håndtere privilegerede konti og skabe udførlige session revisions-logs og videooptagelser af alle privilegerede it-administrator sessions på remote enheder.	Nej	AP understøtter ikke RPA.	
13.12	Projektet skal vurdere, hvor det kan give mening at implementere sessionsstyrings-kapabiliteter som fx skærmbilleder eller optagelser af handlinger, for både at forebygge og fange utilsigtet brugeradfærd fra robotter og de menneskelige brugere bag.	Nej	AP understøtter ikke RPA.	

13.13 (kun relevant ved kritiske systemer)	Det skal sikres, at eventuelle RPA-værktøjer genererer en komplet, system-genereret log der ville kunne understøtte en eventuel revision, eller fremtidigt efterforsknings-arbejde. RPA logs skal lagres på et separat system, beskyttet af kryptering, adgangsstyring og monitorering, således at revisionssporet er intakt og troværdigt. Det er projektets ansvar at være i overensstemmelse med koncernens retningslinjer for disse områder.	Nej	AP understøtter ikke RPA.	

10.2 Procedure for idriftsættelse af model på Analyseplatformen

10.2.1 Step

10.2.2 Indhold

10.2.3 Ansvar

Før idriftsættelse

Review af model/kode:

Udviklerteam laver eget review hele vejen igennem opgaven, inkl. opfølgning på ReFors-tjeklisten - se her ReFors-tjekliste[36]. Hvis man er alene på en opgave, har man pligt til at finde en Data Scientist-kollega, som man løbende kan sparre med og som løbende kan være med til at kvalitetsikre modellen.

To til tre uger, inden modellen lægges i produktion, skal der foretages et eksternt review. Ekstern reviewer (Data Scientist) laver et 2 timers high-level review inden idriftsættelse. Her gennemgås ReFors-tjeklisten (udfyldt), og der laves stikprøver.

Efter reviewet har udvikleren så to-tre uger til at udbedre de eventuelle kritiske fejl og mangler, der blev identificeret i reviewet.

Efter internt review og en korrekt udfyldt ReFors-tjekliste, samt et godkendt eksternt review, er modellen klar til idriftsættelse hos Avanceret Analyses MLE-team. Idriftsættelsen følger de retningslinjer, der er besluttet ift. sikkerhed og kvalitet; se herunder.

36 https://confluence.ccta.dk/display/daiufst/ReFors+tjekliste

Udviklerteam/Data Scientists, fx i DA eller PA

Idriftsættelse

Når ovenstående er gennemført, kontaktes MLE-teamet mhp. idriftsættelse af modellen på Analyseplatformen.

Forud for idriftsættelsen skal MLE sikre, at ”formalia” er overholdt:

- Sikkerhedsgodkendelse
- Refors-tjekliste udfyldt
- Dokumentation er OK
- Data-kilder, data-modtagere, model-ejerskab er defineret.

MLE sikrer, at:

- Koden følger automatiske regler og overordnet flow
- Koden ser fin ud (stikprøver)
- Git er benyttet fornuftigt i udviklingen.

Etablering af overvågning??

Når ovenstående er gennemført, deployes modellen til produktion.

MLE-teamet

I drift

(Kan I skrive noget her? - hvad har MLE/andre ansvar for??)

??

11 Information om adgange og adgangskontrol

11.1 Typer af adgange

På Analyseplatformen har man behov for en række adgange. Disse adgange varierer i forhold til hvordan man bestiller dem, og hvem der har ansvaret for dem.

11.1.1 Omada-adgange

Systemadgange, der bestilles via Omada. Det er din Omada-ansvarlige (typisk kontorchef) der skal bestille adgangen til dig.

Linket til Omada er her https://iam.ccta.dk/dashboard.aspx[37]

For de teknisk interesserede: En adgang i Omada er knyttet til en AD gruppe (en gruppe i SKATs Active Directory), og disse AD grupper kan dermed bruges overalt i SKAT til at styre adgange. Disse gruppemedlemskaber kan enhver også slå op selv. Se mere om det længere nede. Omada-adgange bruges bredt i hele SKAT. Så der er både Omada-adgange som vi er direkte ansvarlige for på Analyseplatformen og der er også rettigheder, styres af andre i SKAT, f.eks. DevOps kontoret eller IT.

11.1.2 Adgange til datawarehouse

Hvis man skal tilgå data i SKATs Sybase Datawarehouse, så skal man dels have oprettet en bruger og dels have tildelt adgang til de specifikke tabeller. KDW har indført, at der først skal søges adgang gennem Omada[38] adgangen hedder "Direkte tabeladgang til Koncern Data Warehouse (KDW)", først herefter skal der søges adgang til den individuelle tabel gennem MitIT som linkes nedenfor. Disse adgange administreres direkte i Sybase, og det er derfor Datawarehouse-gruppen der styrer det og ikke hverken IT eller os på Analyseplatformen. Denne adgang til at hente data fra Datawarehouse er også en anden adgang end adgange til BO og QLIK. De er helt adskilte. Så man kan ikke tilgå en tabel fra R bare fordi man kan tilgå den i BO - eller omvendt.

Brugeroprettelse og adgange bestilles via denne form[39] på MitIT. Anmodningen ender så hos kontoret Data Warehouse og Business Intelligence - så det er altså ikke noget, som vi kan gøre på Analyseplatformen.

Hvis en brugers password er udløbet kan de tilgå dette link og løse det https://dw-password-changer.rke-prod-01.k8s.internal

37 https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fiam.ccta.dk%2Fdashboard.aspx&data=05%7C01%7CPer.Hansen%40ufst.dk%7Ca768e4c89e5f410985de08da599dda84%7C2e93f0edff3646d49ce6e0d902050cf5%7C0%7C0%7C637920831747543267%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=suyEAEvtuGR5746sI1Y5Wv4tmaDQiksNpdc5Cq2A3YU%3D&reserved=0

38 https://iam.ccta.dk/workitemdlg.aspx?ACTTEMP=1000720&RURLID=aff2823c-80f5-4ff0-b345-5a9c1b81987e

39 https://skat-myit.onbmc.com/dwp/rest/share/OJSXG33VOJRWKVDZOBST2U2SIQTHIZLOMFXHISLEHUYDAMBQGAYDAMBQGAYDAMJQGMTHEZLTN52XEY3FJFSD2U2SI5AUCNKWGBDDKM2CGFAVARRZGFDUWUCFJVKFGN2RIFHVEJTDN5XHIZLYORKHS4DFHVBUCVCBJRHUOX2IJ5GUKJTQOJXXM2LEMVZFG33VOJRWKTTBNVST243SNU======

11.2 Liste over relevante Omada adgange på Analyseplatformen

11.2.1 R Studio

- **Giver adgang til:** At logge på R Studio serveren og afvikle R kode.
- **Findes gennem:** Omada[40] ved at søge på det team/selfserviceprojekt du skal være en del af. I Center for Avanceret Analyse har vi fx 'Takamaka Udviklerteam' og 'Atlantis Udvikler team', men hvis et projekt har særskilt adgangsstyring skal du også søge denne gennem Omada .
- **Tilhørende AD gruppe:** ap_r_user_analyse og ap_r_user_skat
- **Note:** Der er 2 forskellige R Studio servere. En til personer i Avanceret Analyse - R03: https://ap-rstudio1.ccta.dk - og en til resten af SKAT - R02: https://ap-rstudio2.ccta.dk.

11.2.2 Jenkins

- **Giver adgang til:** At logge på Analyseplatformens Jenkins server: https://ap-jenkins.ccta.dk/ hvor man kan se Analyseplatformens produktionspipeline for alle projekter.
- **Tildeles:** automatisk på baggrund af anmodningen til R-studio.
- **Tilhørende AD gruppe:** ap_jen_user

11.2.3 Shiny-apps

- **Giver adgang til:** De enkelte produktionslagte Shiny-applikationer på Analyseplatformen - dvs. brugeradgang ikke udvikler-adgang.
- **Findes under:** Analyseplatform APPs
- **Tilhørende AD gruppe:** Separat AD gruppe til de enkelte applikationer. Alle startende med "ap_shiny_"/"ap_user_"

11.2.4 Github

- **Giver adgang til:** At logge på SKATs fælles Github server: https://github.ccta.dk/.
- **Findes under:** GitHub
- **Tilhørende AD gruppe:** github_searchbase
- **Info:** Github serveren administreres IKKE af os (Analyseplatformen) men af DevOps. Men hvis man vil udvikle Python pakker på Analyseplatformen, så skal de lægges på Github, fordi vi benytter DevOps's pipeline til at bygge Python pakker.
  Skriv dit w-nummer med lille w.

40 https://iam.ccta.dk/dashboard.aspx

11.2.5 SAS

- SAS miljøet er ikke længere en del af Analyseplatformen, men administreres af Produktionssatte Dataløsninger. Du kan finde information om SAS platformen på deres Confluence side: FAQ for SAS DataAnalyse Brugere[41]
- **Findes under:** Analyseplatform SAS og SAS VA
- **Tilhørende AD gruppe**: ap_sas_user

11.3 Check adgange og gruppemedlemskaber i AD

Alle Omada-adgange er knyttet til en AD gruppe, og man kan selv undersøge hvilke grupper man selv eller andre er medlem af eller hvem der er medlem af en given AD gruppe.

11.3.1 Check adgange vha. grafisk tool

Der findes en lang række værktøjer til at browse Active Directory. F.eks. Microsofts egen ADExplorer. Bare forbind til serveren hocdc01.ccta.dk og login med dine Windows credentials (wnr).

11.3.2 Check adgange fra en Linux kommando-prompt med ldapsearch

Fra en Linux kommandolinie kan man bruge ldapsearch til at hive informationer ud fra AD. For at snakke med AD serveren, så skal man først lave en kerberos ticket med kommandoen "kinit".

Derefter kan man forbinde til en AD server og lave forespørgsler. Hvis du vil kende detaljerne, så kig i kildekorden for de scripts, der beskrives under Hjælpeværktøjer på R Studio serverne[42]

Blandt hjælpeværktøjerne, der er der 2 specifikke kommandoer, til at checke rettigheder i AD:

For at finde alle dine eller andre brugeres gruppemedlemskaber kan du benytte kommandoen getmemberof. F.eks.
getmemberof w19683
For at se alle de grupper, som den bruger er medlem af.

For at finde alle medlemmer af en specifik grupper, så er det smartest at benytte den indbyggede kommando getent group, og der er lavet meget lille script "ggg" som gør det. F.eks.
ggg ap_r_user_analyse
For at se wnr for alle medlemmer i gruppen ap_r_user_analyse

11.3.3 Check adgange fra en Windows kommando-prompt:

Man kan gøre det samme i Windows

For at se information om en bruger - inklusiv gruppe-medlemskaber:
net user w19683 /domain

41 https://confluence.ccta.dk/display/PD/FAQ+for+SAS+DataAnalyse+Brugere
42 https://confluence.ccta.dk/pages/viewpage.action?pageId=378465566

For at se alle medlemmer af en specifik grupper:
net group ap_r_user_analyse /domain

12 Udvikling vs. Produktionslægning

12.1 Manglende R pakker

Udviklere oplever ofte, at de ikke har adgang til en pakke, som de har behov for. R-pakker bliver tildelt de enkelte kontorer efter behov. Hvis man har yderligere behov for pakker, så kan man anmode om adgang til yderligere pakker
i MitIT.

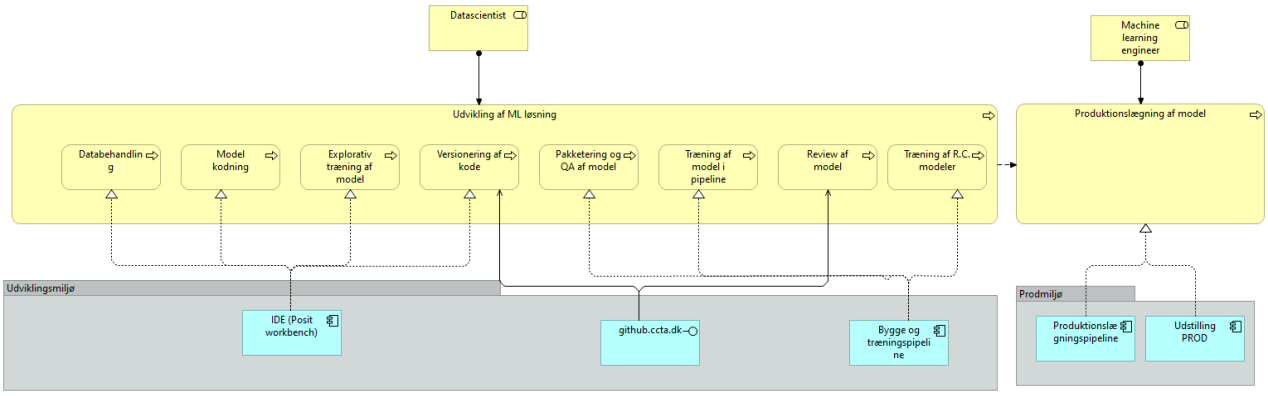
Det er typisk adgang til pakker til "avanceret analyse", som man ikke har adgang til. Det er ikke noget, som vi gør for at genere eller for at forhindre folk i at udføre deres arbejde. Grunden til at vi gør det, er fordi vi godt vil sørge for, at der ikke udføres analyse-opgaver i isolerede øer rundt omkring i SKAT, uden at der er styr på detaljerne omkring sporbarhed, dokumentation udviklingsprincipper osv. Så længe man udfører "one-off" skrivebordsanalyser, så er det måske godt nok, men hvis man begynder f.eks. at køre periodiske rapporter eller - endnu vigtigere - hvis de analyser/modeller der udvikles på nogen som helst måde påvirker borgere (via. prioritering, sagsbehandling, nye regler eller andre afledte effekter), så SKAL tingene lægges i produktion på den korrekte måde.

12.2 Analyseplatformen

Vi har udviklet Analyseplatformen til at hjælpe udviklere i SKAT med at lave Machine Learning modeller m.m. og produktionslægge modellerne på helt struktureret vis.

Produktionslægning sikrer mange væsentlige elementer:

- Overblik over hvilke modeller/analyser/rapporter der kører i produktion
- Historik og versionskontrol ved at koden lægges ind i Git
- Dependency control, så koden ikke pludselig holder op med at virke pga. software-opdateringer.
- Ensartede metoder / udviklingsprincipper som sikrer både kvalitet, mulighed for review og en meget nemmere overlevering til andre.
- Sikkerhed omkring dataadgange og data opbevaring
- En høj grad af automatisering
- En ansartet måde at udstille de udviklede løsninger på.
- Logning
- ... og meget mere.



Pointen er, at der IKKE skal køre modeller i "produktion" på en udviklingsplatform - som R Studio platformen er.

Samlet set, så betyder det, at man skal overveje, om det man sidder med faktisk hører til i produktion.

Hvis det hører til i produktion, så sørger vi for at oprette projektet i vores produktionspipeline. Det betyder bl.a. at man får et projekt på vores Git-server, og at man skal udvikle sin kode efter nogle standarder og skabeloner, som vi har udviklet. Det betyder f.eks. at alt hvad man laver, bliver lavet som selvstændige R-pakker.

Der er flere måder det kan foregå på:

1. Hvis man har mod på det, så kan man fint gøre det helt selv. Der er nogle ting man skal være opmærksom på og skal lære, men der er ikke noget i vejen for, at man helt selv udvikler en pakke og får den produktionslagt.
   Vi hjælper gerne med spørgsmål, og man kan finde en del svar i vores FAQ i Confluence: https://confluence.ccta.dk/display/daiufst/FAQ+-+Udvikling+i+R+Studio
2. Vi kan også hjælpe med udviklingen. Vi kan finde en Data Scientist i Avanceret Analyse, som kan assistere med udviklingen eller som man kan spare med i det omfang, det er nødvendigt. Det kan evt. være en god måde at blive introduceret for Analyseplatformen og vores udviklingsprincipper.
3. Man kan få os til at udvikle en model/analyse hvor man selv primært bidrager med forretningsviden. Det betyder at opgaven skal bestilles via de normale procedurer og skal igennem en prioritering og bliver tildelt et af de teams, der sidder i Avanceret Analyse.

Det er klart, at det **VIL** være mere omstændigt, end bare at sidde og adhoc-kode derudaf i R Studio, men det er nødvendigt, at produktionsopgaver ikke flyder rundt på mærkelige steder i Skatteforvaltningen. Dels fordi det er ulovligt, men også fordi vi skyder os selv i foden på længere sigt.

Så hvis du sidder med udviklingsopgaver, som ligner noget, der burde lægges i produktion, så tag fat i os. Opret en sag i MitIT eller skriv en mail til de MLE'ere, der udvikler og driver Analyseplatformen på analyseplatform@ufst.dk.[43]

43 mailto:analyseplatform@ufst.dk.