

Promilleafgiftsfonden for landbrug

Notat

SEGES Innovation
Digital

Afgrøde prognose - opsamling	Ansvarlig	HJN
	Oprettet	30-08-2023
Projekt: 2555:2315 – Succes med vandmiljøet	Side	1 af 41

Formål

Dette notat er en kort opsamling af arbejdet med fremskrivning af afgrøde på markniveau, statistikrapporten og dataforklaringsnotatet.

Yderligere beskrivelser og resultater, der danner baggrund for denne opsamling, findes i bilagene.

Bilag 1. Statistikrapport

Bilag 2. Databeskrivelse

Opgaven var "Hvor godt kan vi gætte næste års afgrøde?".

Den stillede opgave var at generere og rense data til modelinput. Afprøve en model til hvor godt vi kan prædiktere næste års afgrødevalg på markniveau ud fra markens historik. Samt se på de bedriftsspecifikke output fra modellen.

Data

Der er arbejdet med at generere et fuldstændigt datagrundlag, hvor der blandt andet har været fokus på at koble markhistorik til at blive bedriftshistorik. Datagrundlaget der danner udgangspunkt, er dyrkningsoplysninger fra ansøgning om grundbetaling år for år i perioden 2011-2022. Der er selvsagt ikke alle marker der drives af samme bedrift hvert eneste år, men modelmæssigt er det vigtigt at have styr på hvilke marker hver bedrift rådede over i det enkelte år, og at sammenstille data, så det er muligt at kortlægge dyrkningshistorikken på markerne, som bedriften råder over.

Der har været et arbejde med at definere afgrødegrupper, definere bedriftstyper, se på sekvensmønstre i forhold til frie år mellem afgrøder, samt indledende dataundersøgelser. Derudover har der været brugt tid på at sammenstille modellens output med de bedriftsspecifikke data for at præsentere data ned på bedriftsniveau, der har været anvendt til blandt andet at samle og omdefinere afgrødegrupper.

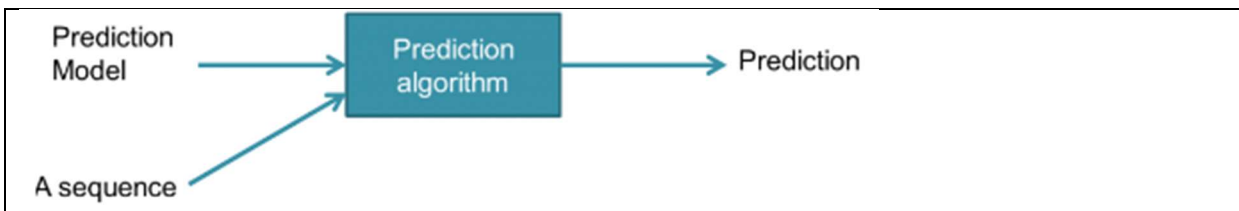
Model

Tilgangen for at løse en sådan problemstilling er som nedenstående.

Første skridt til en sådan sekvens prædiktion er at generere træningssekvenser, træne en model og dernæst ende med en prædiktionsmodel.



Andet skridt er derefter at bruge denne trænedede sekvensprædiktionsmodel til at prædiktere udfaldet af nye ukendte sekvenser.



Der har været skimmet forskellige modelteorier for at kunne håndtere det der kaldes "sequence prediction problems", og som her er den type sekvens prædiktions som kan kaldes "many-to-one". Der har været kig på LSTM-modeller, RNN-modeller, XGBoost, Random Forest med tidsafhængige prædiktorer og alle med tanke på et stærkt ubalanceret datasæt.

Det er ikke en tidsserie problemstilling, da vi ikke har numeriske variable, men kategoriske afgrødegruppesekvenser.

I starten var der indledende undersøgelse af RNN-modeller, da disse lyder lovende til at håndtere problemstillingen med tidsafhængige sekvenser. Der opstod tekniske udfordringer med at få afviklet modellen i R, både i forhold til selve teknikken, samt opbygningen af input data. Så denne blev nedprioriteret og der blev forsøgt med en mere simpel tilgang, hvor modelafviklingen er mere simpel og også computertidsmæssigt hurtig.

Valget blev at teste om en XGBoost model ville kunne give en tilstrækkelig god præcision til en start. Fordelene ved en XGBoost er at denne model er stærk i mønster-genkendelse, og den er rent teknisk hurtig at afvikle. Modellen er sammenlignet med f.eks. en randomforest model også bedre til at håndtere ubalancerede data.

Ulemperne kan dog være at der er mange komplekse parametre i modellen, der kan tunes for at give bedste resultat og ligeledes var det ikke givet at sekvensmønstrene blev fanget godt nok.

Modelresultater

Ud over kun at se på modellens output parametre i forhold til overordnet præcision og confusions matrix, har det også været ønsket at se ned på de enkelte afgrøder og de enkelte bedrifter (type, region osv). Baggrunden for dette har været at undersøge om en bestemt type bedrifter kunne opnå en meget høj præcision, og der dermed formidlingsmæssigt og output-mæssigt ville kunne fokuseres på denne delmængde af bedrifter.

Dette har der ressourcemæssigt også været prioriteret.

Hovedkonklusioner

Den simple tilgang med en XGBoost model har ikke vist sig at være tilstrækkelig i forhold til præcision. Modellen var ikke god nok til at genkende mønstrene i marksekvenserne, og bruger ikke al historik til at fremskrive næste års afgrødevalg. Dermed udgår væsentlige afgrødegrupper i prædiktionen. Årets udbytte har dermed ikke været fuldstændig som håbet, men der er dog opnået en væsentlig viden om strukturering og klargøring af data til opgaver af denne type. Fremtidig modellering kan i vid udstrækning baseres på det nuværende datagrundlag. Dermed kan fremtidigt arbejde inden for dette felt, få større fokus på modelegenskaber og modelvalg, da den grundlæggende datahåndtering er på plads.

Fremtidige perspektiver

Der må forventes at være forbedringsmuligheder ved modeller der håndterer tidsafhængigheden i sekvenserne bedre. Nogle hvor der er større fokus på det historiske perspektiv end blot foregående 1-2 års afgrøder (her med tanke på eksempelvis raps og kartofler der har flere frie år imellem).

Et naturligt næste skridt, kunne være at se på Compact prediction tree modeller, som er en senere udvidelse af RNN-modeller. Disse har blandt andet fordele i at være hurtigere. Og tyder på at der heller ikke skal gentrænes ved usete afgrødegrupper. Yderligere undersøgelser af modelteori vil indgå i fremtidige projekter.

Bilag 1. Statistikrapport

STATISTIKRAPPORT

TITEL: Afgrøde prognose - Modelrapport XGBoost
PROJEKT 2255 - Succes med vandmiljøet
STATISTIKER: Heidi Lund Hyttel <hjn@SEGES.dk>
DATO: 2023-08-10

Formål

Formålet med dette notat er at beskrive den første iteration af modeludvikling for at prædikere næste års afgrøde valg. Der er ligeledes i rapporten koblet outdata til bedriftens karakteristika. På den måde gives et overblik over hele bedriften og hvor meget der rammes af bedriftens afgrødevalg.

Data overblik

Se appendix for dataforklaring.

Nedenfor ses forskellige optællinger og fordelinger af datainput på eksempelvis afgrøde-gruppe, region, økologi, bedriftstype, jordbundstype. Ligeledes ses fordelingen af antal marker per bedrift.

Tabel 1: Fordeling af bedrifter, marker og ha

Antal.Bedrifter	Antal.Marker	Antal.ha.2021
24.793	366.051	1.856.346

Tabel 2: Fordeling af afgrødegrupper

GrpNavn	Antal marker	Antal ha 2021
PermanentGraes	86.620	163.174
Vaarsaed	81.970	507.973
Vintersaed	68.141	546.688
Fodergraes	50.651	188.715
ArealUdenProd	29.135	32.338
Majs	19.787	128.072
Vinterraps	13.375	126.613
Frograes	7.227	82.032
Kartofler	3.443	25.256
Restgruppe1	3.169	26.182
Sukkerroer	2.287	27.966
Havefrø	160	732
Restgruppe2	44	228
OliefroeOvr	40	371
VinterkornHelsaedGroenkorn	2	6

Tabel 3: Fordeling af Region

Region	Antal marker	Antal ha 2021
Region Midtjylland	114.437	548.459
Region Syddanmark	108.984	544.941
Region Nordjylland	74.892	353.359
Region Sjælland	52.425	341.845
Region Hovedstaden	15.303	67.565

Tabel 4: Fordeling af Økologi

Okø_2021	Antal marker	Antal ha 2021
0	325.802	1.682.319
1	40.249	174.028

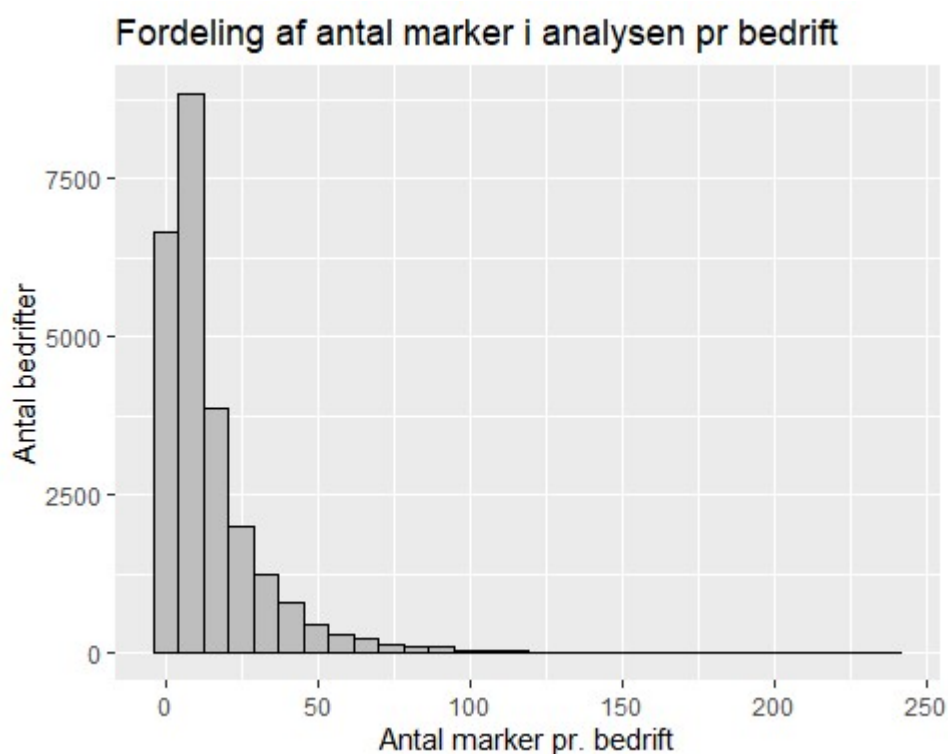
Bedriftstyperne defineres som "Froe" hvis arealandelen af frø ("Frograes", "Spinat", "Havefrø") er mere end 15 %, "Kvaeg" hvis arealandelen af kvaeg ("Fodergraes", "Majs") er mere end 50 %, "Vinter" hvis arealandelen af vinter ("Vinter") er mere end 50 % og ellers "Other".

Tabel 5: Fordeling af Bedriftstype

BedriftType	Antal marker	Antal ha 2021
Other	191.428	809.236
Vinter	95.530	591.483
Kvaeg	55.614	276.508
Froe	23.479	179.120

Tabel 6: Fordeling af Jordbundstype

JB_gruppe	Antal marker	Antal ha 2021
JB_5_9	126.109	788.172
JB_1_3	102.330	464.546
JB_2_4_10	101.659	501.133
JB_11	35.714	101.738
0	239	758



Figur 1: Fordeling af antal marker pr bedrift

XGBoost

Der er i denne analyse valgt at afprøve modellen XGBoost "Extreme Gradient Boosting" som en indgang til opgaven.

Denne model er en træbaseret machine-learning algoritme som kort fortalt kører nogle serielle træstrukturer og derved lærer af de tidligere seriers fejl for at forbedre dem og dermed øge præcisionen af modellen.

En af fordelene ved at teste XGBoost er også performance. Det går rigtig hurtigt når modellen skal trænes og prediktere. Med et forholdsvis stort datasæt som dette, da er det også en vigtig

parameter. Ved at bruge en XGBoost model frem for eksempelvis en Random Forest model, da vil denne model også bedre kunne håndtere den ubalance vi har i data, i og med nogle afgrødegrupper er rigtig store og nogle rigtig små.

For yderligere at håndtere denne ubalance i afgrødegrupperne, har vi blandt andet vægtes med. Her vægtes der i modellen med antal marker i afgrødegruppen for ikke at overprædiktere de små afgrødegrupper. Dette gør dog så også at nogle af afgrødegrupperne er så små at der slet ikke prædikteres for dem.

Ved test af modellen er der først testet 7, 15 og dernæst 30 iterationer som det maximale antal boosting iterationer. Der er ved afvikling af modellen observeret at ved 30 iterationer bliver error-rate for modellen ikke betydelig mindre. Der er ikke tunet yderligere på denne parameter med andre metoder, men blot ved observation valgt at fortsætte med max 30 iterationer. I modellen er brugt det der hedder "objective = multi:softmax", som er en metode der bruger multiklassifikation som modevaluering.

Med i modellen er blandt andet også JBgrp samt markens økologistatus

Modelresultater

Hovedkonklusioner

Med en Kappa-værdi på 0.61 for 2021 prædiktionen, da er den samlede model i stand til at give en fremskrivning med en overordnet præcision på ca. 61 procent. Det er moderat, men dog ikke overbevisende som grundlag for en fremskrivning.

Gruppen "vårsæd" har en F1-værdi på 0.55 hvilket er et udtryk for både modellens evne til at fange alle de rigtige vårsædsafgrøder (recall/sensitivity = 0.7), men også et udtryk for hvor præcist vårsæd fremskrives. Det vil sige hvor rigtige er dem der er prædikteret som vårsædsafgrøder (precision = 0.46).

Vintersæd er en anelse mere præcis med F1 = 0.57, mens majs kun opnår F1 = 0.49. De afgrødegrupper som bedst fremskrives, er fodergræs (F1 = 0.78), permanent græs (F1 = 0.78) og arealer uden produktion (F1 = 0.91).

Man kunne med disse resultaterne tænke at modellen har særlig vægt på observationen der kommer i året forud for prædiktionen, og ikke ret meget på afgrødehistorikken i de foregående år. Dette kan være medvirkende til at permanent græs, der i sagens natur ofte ligger samme sted år efter år, er særdeles godt prædikteret, hvilket også gælder for fodergræs og arealer uden produktion. Desværre evner modellen ikke at ramme arealer med frøgræs, og den er ikke ret imponerende til at forudse valg af majs, vårsæd og vintersæd. Raps er ikke med i nedenstående tabel, hvilket skyldes at modellen desværre ikke kan forudse raps som afgrøde, til trods for at der er samme areal som majs, men fordelt på væsentligt færre marker.

Tabel 7: Kappa og F1 metrikker for 2021 og 2022 prædiktionen

Prædiktion	Kappa	F1 - Vaarsaed	F1 - Frograes	F1 - Vintersaed	F1 - Majs	F1 - Permanent-Graes	F1 - Foder-graes	F1 - ArealUden-Prod
2021	0,61	0,55	0	0,57	0,49	0,96	0,78	0,91
2022	0,16	0,48			0,00	0,01	0,30	0,64

Tabel 8: Alle metrikker for 2021 prædiktionen

GrpNavn	Sensitivity	Specificity	Precision	Recall	F1	AntalMarker_2021
Vaarsaed	0,69538856	0,7648030	0,4603686	0,69538856	0,5539832642	81970
Frograes	0,00013837	1,0000000	1,0000000	0,00013837	0,0002767017	7227
Vintersaed	0,54864179	0,9159310	0,5988307	0,54864179	0,5726386410	68141
Kartofler	0,00000000	1,0000000		0,00000000		3443
Sukkerroer	0,00000000	1,0000000		0,00000000		2287
OliefroeOvr	0,00000000	1,0000000		0,00000000		40
Majs	0,34942134	0,9949056	0,7967274	0,34942134	0,4857895661	19787
Vinterraps	0,00000000	1,0000000		0,00000000		13375
PermanentGraes	0,96660125	0,9832517	0,9470630	0,96660125	0,9567323899	86620
Fodergraes	0,81291584	0,9551078	0,7441175	0,81291584	0,7769967448	50651
ArealUdenProd	0,88151708	0,9949483	0,9378492	0,88151708	0,9088110403	29135
Restgruppe1	0,00000000	1,0000000		0,00000000		3169
Restgruppe2	0,00000000	1,0000000		0,00000000		44
Havefrø	0,00000000	1,0000000		0,00000000		160
VinterkornHelsaedGroenkorn	0,00000000	1,0000000		0,00000000		2

Mål for model performance

I nedenstående ses betydningen af de forskellige modelmetrikker der tales om i dokumentet her.

Tabel 9: Betydning af modelmetrikker

Metrik	Betydning
Sensitivity	Andelen af de afgrødegrupper 'X' der er rigtigt prædikeret som 'X'
Specificity	Andelen af de afgrødegrupper der er 'ikke X' som prædikeres rigtigt som 'ikke X'
Precision	Andelen af afgrødegrupper der er rigtigt prædikeret som 'X', ud af alle dem der er prædikeret som afgrødegruppe 'X'
Recall	Samme som Sensitivity
F1	Harmonisk gennemsnit mellem precision og recall. Modellens evne til at både at fange de positive tilfælde (recall) og prædikere rigtigt på de positive prædiktioner (precision). F1 er god som metrik for ubalanceret data som vi har her. Man kan som tommelfingerregel sige (<0.5: Ikke godt), (0.5-0.8: Ok), (0.8-0.9: Godt) og (>0.9: Meget godt)
Kappa	Hvor god er modellen er til at klassificere rigtigt. I forhold til metrikken Accuracy, så tager Kappa-metrikken også højde for den ubalancerede klassefordeling. Man kan som tommelfingerregel sige (<0: Ikke god), (0-0.20: Lidt), (0.21-0.40: Fair), (0.41-0.60: Moderat), (0.61-0.80: Betydelig) og (0.81-1.0: Perfekt)

Model

Modellen trænes på alle data 2011-2020. Herefter bruges denne model til at prædikere for ukendt 2021-data og validere denne op mod det kendte afgrødevalg for 2021.

Derefter bruges denne 2021-prædiktion ind i nyt 2011-2021 datasæt hvor der så prædikteres 2022-afgrødevalg, samt validere denne op imod vores kendte 2022-afgrødevalg. Prædiktionerne for 2022 data er betydeligt dårligere. Men i og med der er knap så god en præcision på 2021-data og der er flere afgrødegrupper som ikke prædikteres, da vil dette resultat være forventeligt.

I nedenstående tabeller ses både prædiktions mål for de enkelte 2021- og 2022-afgrødeprædiktioner. Samt confusion matrix som viser hvordan de reele afgrødegrupper så er prædikeret.

Table 10: Model measures 2021 predictions

GrpNavn	Sensitivity	Specificity	Precision	Recall	F1	AntalMarker_2021
Vaarsaed	0,69538856	0,7648030	0,4603686	0,69538856	0,5539832642	81970
Frograes	0,00013837	1,0000000	1,0000000	0,00013837	0,0002767017	7227
Vintersaed	0,54864179	0,9159310	0,5988307	0,54864179	0,5726386410	68141
Kartofler	0,00000000	1,0000000		0,00000000		3443
Sukkerroer	0,00000000	1,0000000		0,00000000		2287
OliefroeOvr	0,00000000	1,0000000		0,00000000		40
Majs	0,34942134	0,9949056	0,7967274	0,34942134	0,4857895661	19787
Vinterraps	0,00000000	1,0000000		0,00000000		13375
PermanentGraes	0,96660125	0,9832517	0,9470630	0,96660125	0,9567323899	86620
Fodergraes	0,81291584	0,9551078	0,7441175	0,81291584	0,7769967448	50651
ArealUdenProd	0,88151708	0,9949483	0,9378492	0,88151708	0,9088110403	29135
Restgruppe1	0,00000000	1,0000000		0,00000000		3169
Restgruppe2	0,00000000	1,0000000		0,00000000		44
Havefrø	0,00000000	1,0000000		0,00000000		160
VinterkornHelsaedGroenkorn	0,00000000	1,0000000		0,00000000		2

Tabel 11: Confusion matrix - 2021 predictions

Reference	Vaarsaed	Frograes	Vintersaed	Majs	PermanentGraes	Fodergraes	ArealUdenProd
Vaarsaed	57001	0	13192	1137	3143	7148	349
Frograes	5157	1	1987	1	4	72	5
Vintersaed	28739	0	37385	327	319	1246	125
Kartofler	2825	0	447	51	11	97	12
Sukkerroer	1768	0	389	44	11	72	3
OlifroeOvr	28	0	6	0	0	6	0
Majs	9324	0	949	6914	231	2271	98
Vinterraps	7130	0	5991	5	23	198	28
PermanentGraes	211	0	6	1	83727	2266	409
Fodergraes	7601	0	678	102	431	41175	664
ArealUdenProd	1469	0	797	76	471	639	25683
Restgruppe1	2411	0	556	18	34	142	8
Restgruppe2	41	0	1	0	0	2	0
Havefrø	110	0	46	2	1	0	1
VinterkornHelsaedGroenkorn	1	0	0	0	1	0	0

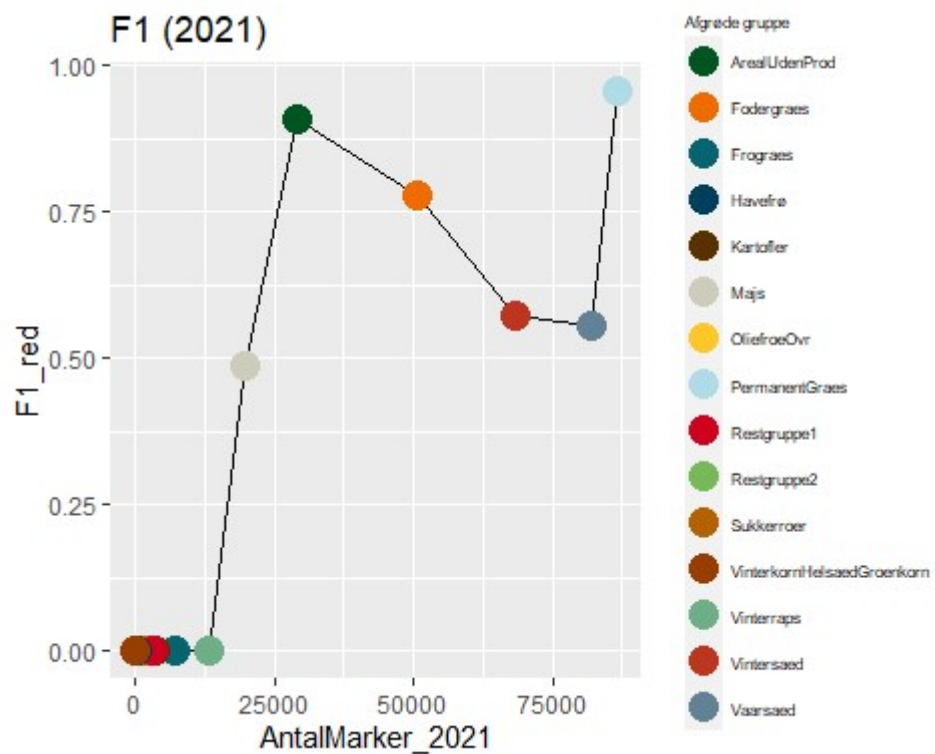
Table 12: Model measures 2022 predictions

GrpNavn	Sensitivity	Specificity	Precision	Recall	F1	AntalMarker_2022
Vaarsaed	0,691212917	0,6508109	0,3618415	0,691212917	0,4750172847	81506
Frograes	0,000000000	1,0000000		0,000000000		7426
Vintersaed	0,000000000	1,0000000		0,000000000		63713
Kartofler	0,000000000	1,0000000		0,000000000		3900
Sukkerroer	0,000000000	1,0000000		0,000000000		2300
OlifroeOvr	0,000000000	1,0000000		0,000000000		33
Majs	0,000359177	0,9999856	0,5833333	0,000359177	0,0007179119	19489
Vinterraps	0,000000000	1,0000000		0,000000000		15530
PermanentGraes	0,004264733	0,9990336	0,5816485	0,004264733	0,0084673813	87696
Fodergraes	0,696726500	0,5387408	0,1891504	0,696726500	0,2975268592	48969
ArealUdenProd	0,623149954	0,9706394	0,6648024	0,623149954	0,6433026433	31283
Restgruppe1	0,000000000	1,0000000		0,000000000		3930
Restgruppe2	0,000000000	1,0000000		0,000000000		49
Havefrø	0,000000000	1,0000000		0,000000000		218
VinterkornHelsaedGroenkorn	0,000000000	1,0000000		0,000000000		9

Tabel 13: Confusion matrix - 2022 predictions

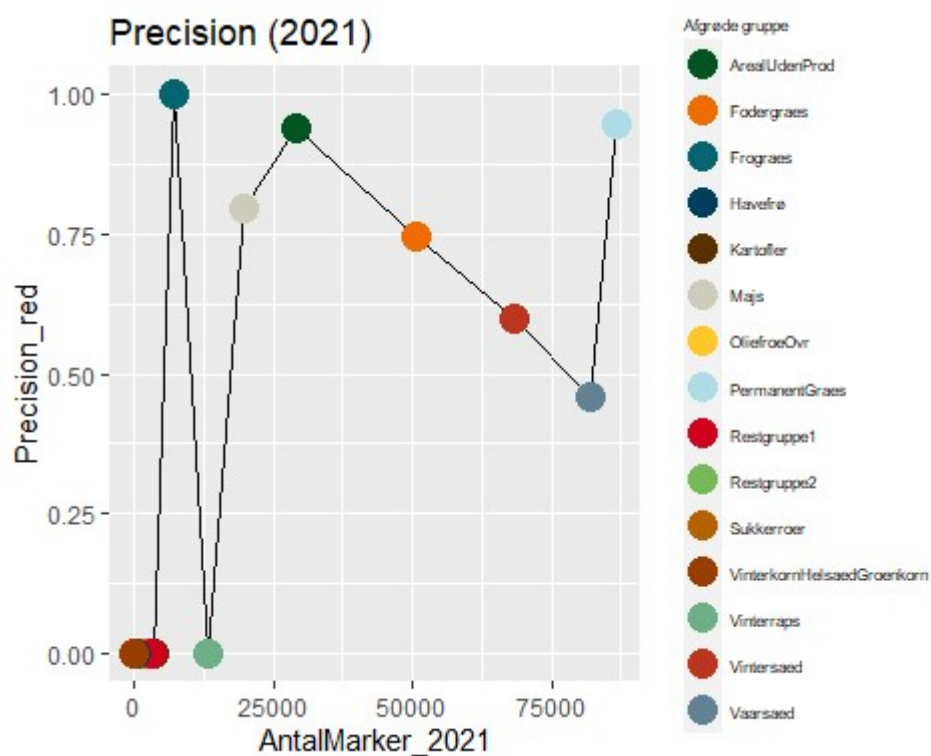
Reference	Vaarsaed	Majs	PermanentGraes	Fodergraes	ArealUdenProd
Vaarsaed	56338	3	74	22188	2903
Frograes	6270	0	0	1066	90
Vintersaed	46255	2	13	13147	4296
Kartofler	3077	0	0	763	60
Sukkerroer	1761	0	0	427	112
OliefroeOvr	20	0	0	11	2
Majs	3909	7	0	15362	211
Vinterraps	11576	0	3	2789	1162
PermanentGraes	3752	0	374	83425	145
Fodergraes	13981	0	147	34118	723
ArealUdenProd	5926	0	21	5842	19494
Restgruppe1	2643	0	11	1158	118
Restgruppe2	19	0	0	29	1
Havefrø	168	0	0	44	6
VinterkornHelsaedGroenkorn	3	0	0	6	0

Nedenfor ses plots af forskellige metrikker fra vores confusion matrix. Metrikkerne er efter ønske sammenholdt med antallet af marker for afgrødegruppen der er med i modellen.



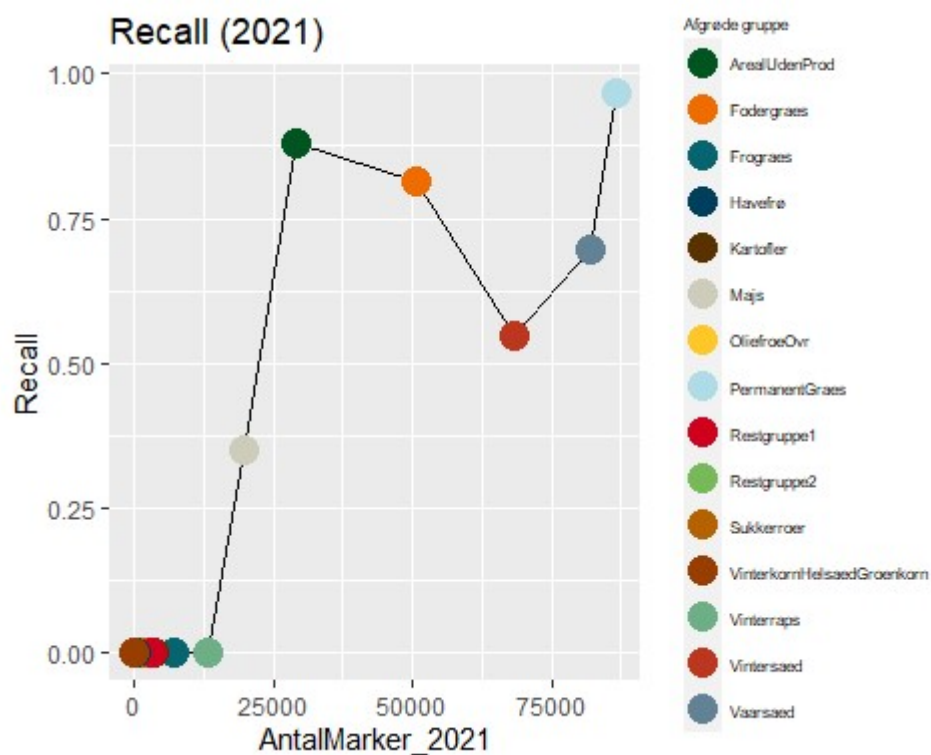
1.

Figur 2: F1 i forhold til antal marker for afgrødegruppen (2021)



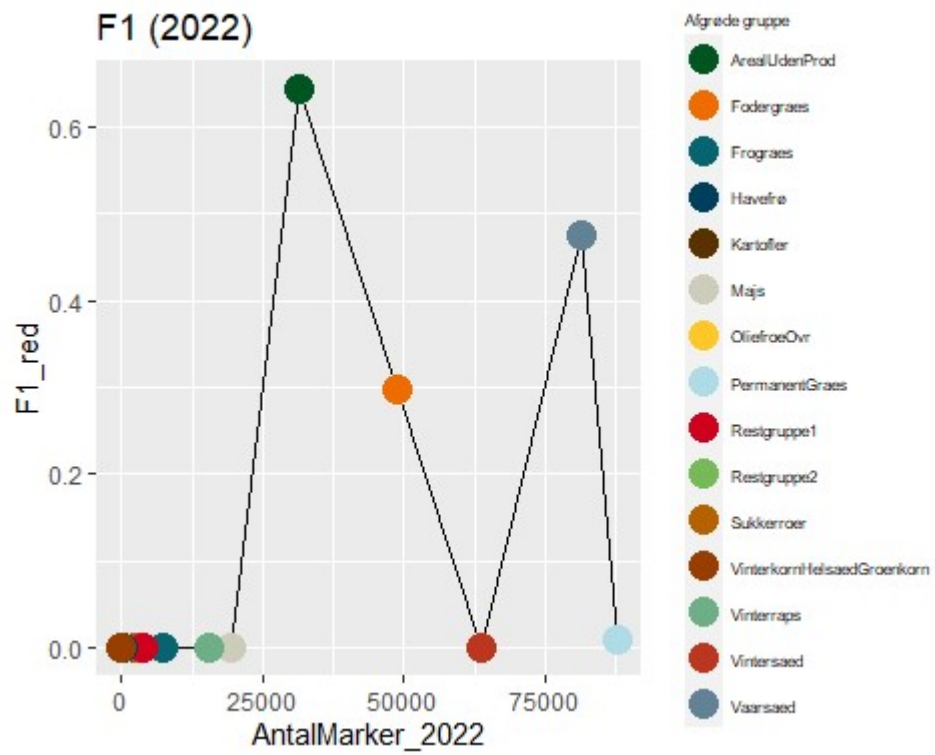
2.

Figur 3: Precision i forhold til antal marker for afgrødegruppen (2021)



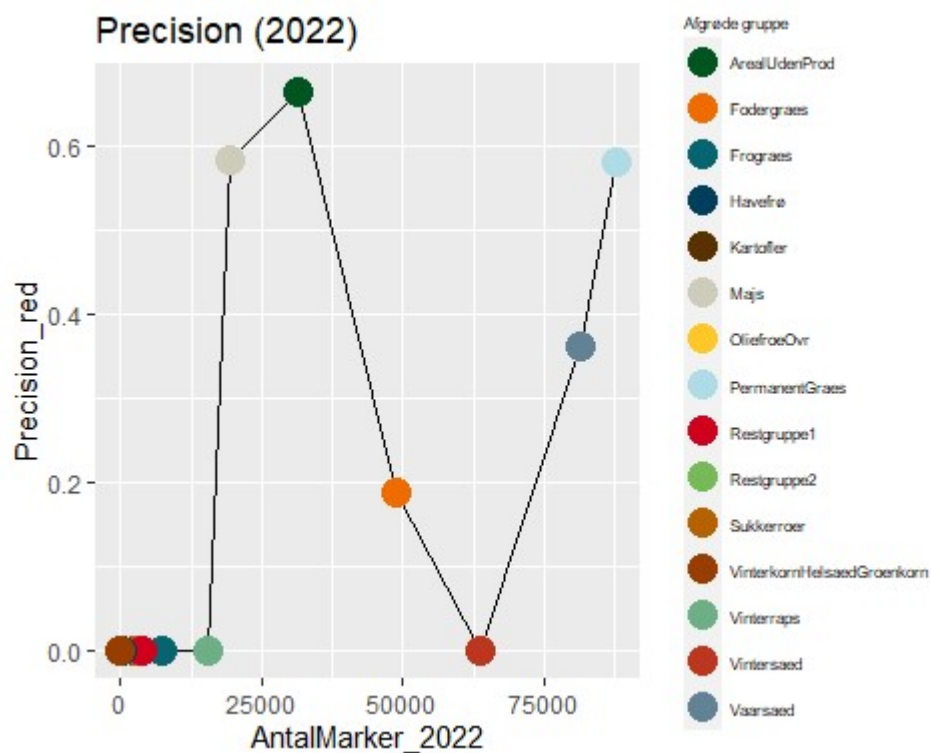
3.

Figur 4: Recall i forhold til antal marker for afgrødegruppen (2021)



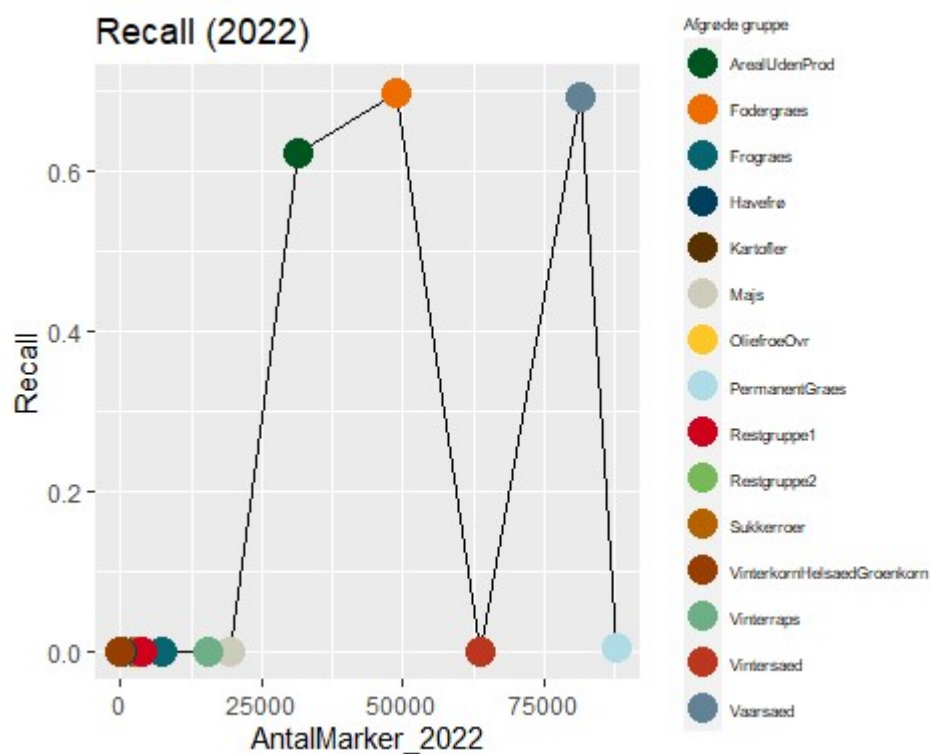
4.

Figur 5: F1 i forhold til antal marker for afgrødegruppen (2022)



5.

Figur 6: Precision i forhold til antal marker for afgrødegruppen (2022)



6.

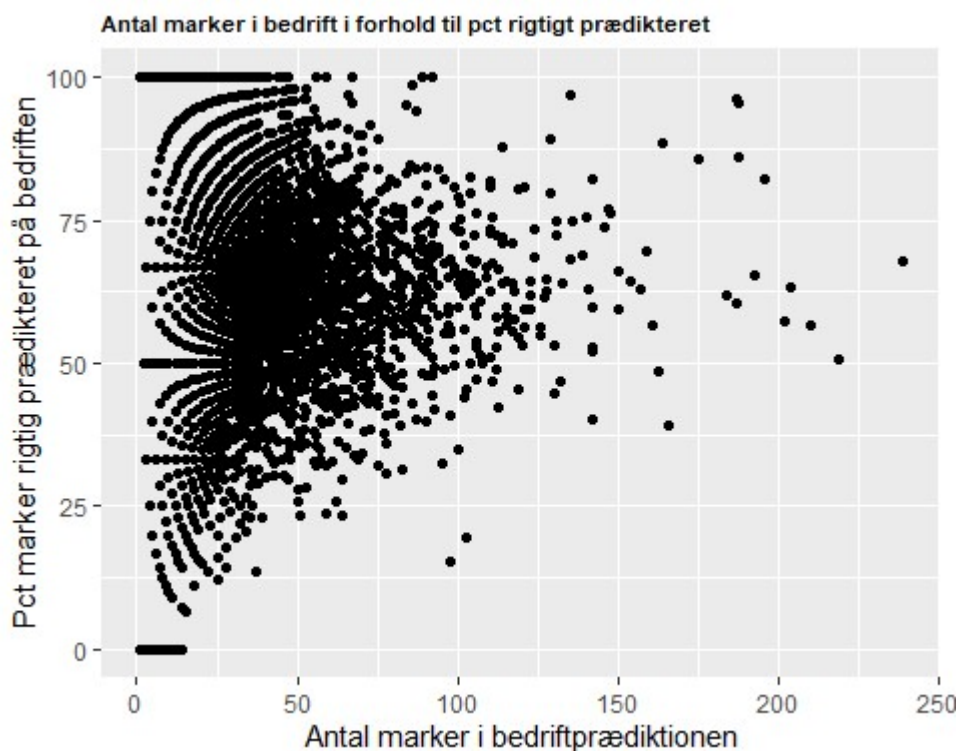
Figur 7: Recall i forhold til antal marker for afgrødegruppen (2022)

Bedriftens resultater

Det er ønsket at finde frem til om der umiddelbart er nogle bedriftskarakteristika der går igen ved de bedrifter hvor der er højest præcision på prædiktionerne. Nedenfor er derfor blandt andet vist bedrifternes "rigtighedsprocent". Det vil sige hvor stor en andel af bedriftens marker er rigtigt prædikeret. Denne er vist i forhold til bedriftstype, areal for 2021, region, økologi, hovedafgrøde.

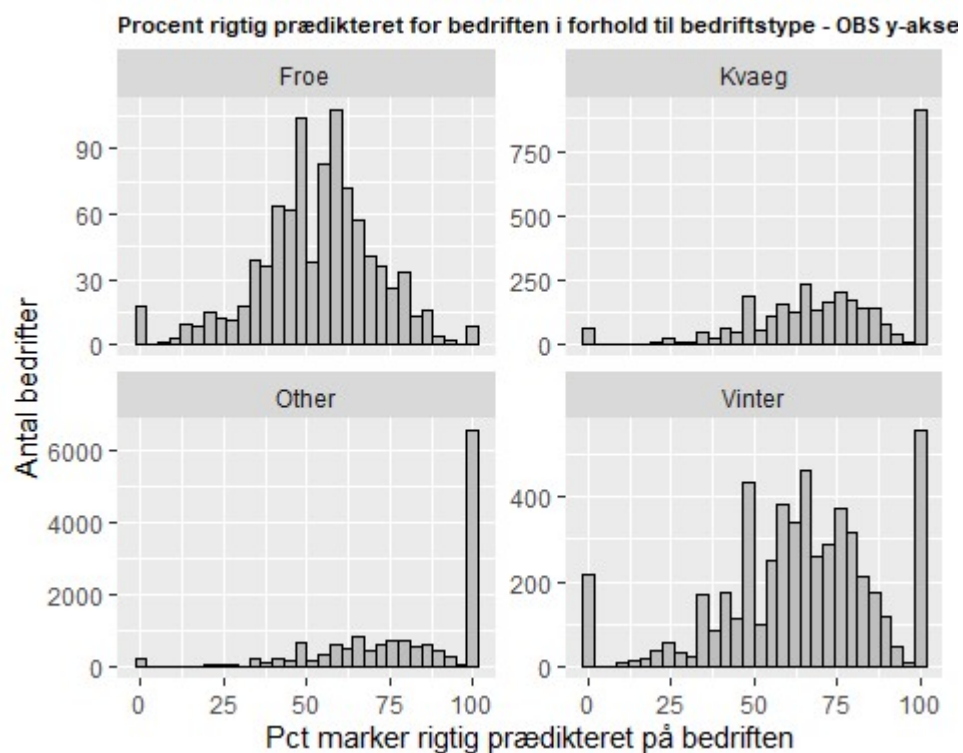
Længere nede kan man dykke mere ned i de bedrifter vi er gode til at prædiktere på. Her er "rigtighedsprocenten" yderligere inddelt i intervaller 0-19 %, 20-39 %, 40-59 %, 60-79 % samt 80-100 %.

I analysen her er der nogle datavalg der gør at der ikke prædikteres for alle bedriftens marker. I appendix kan se hvilke datakrav der er for at vi prædikterer for en given mark.



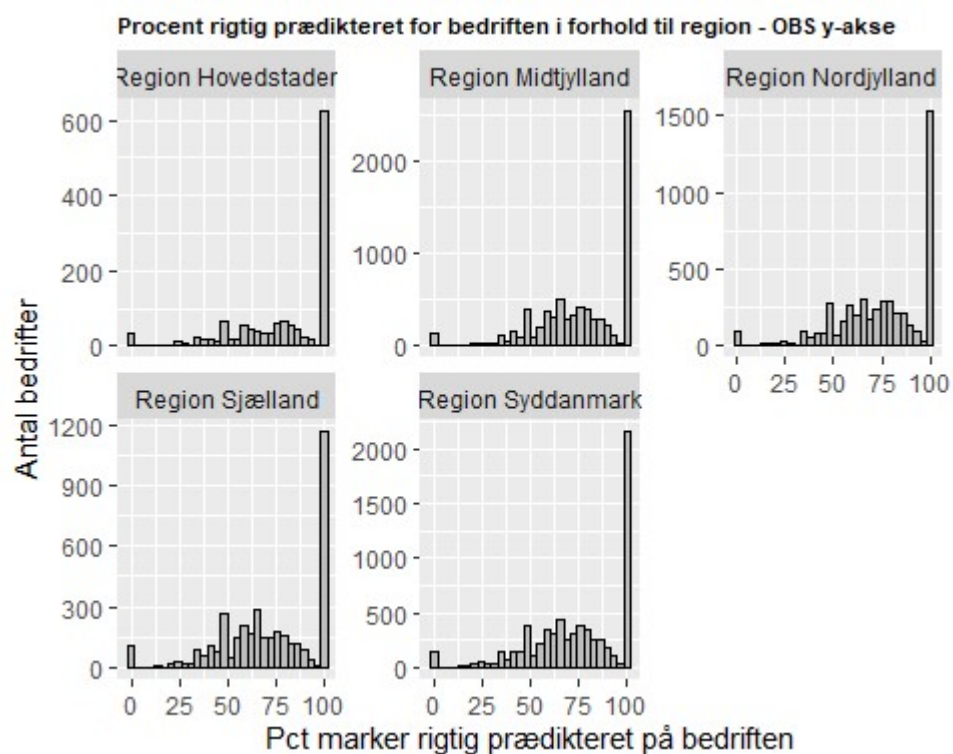
7.

Figur 8: Procent rigtigt prædikeret for bedriften i forhold til antal marker for bedriften



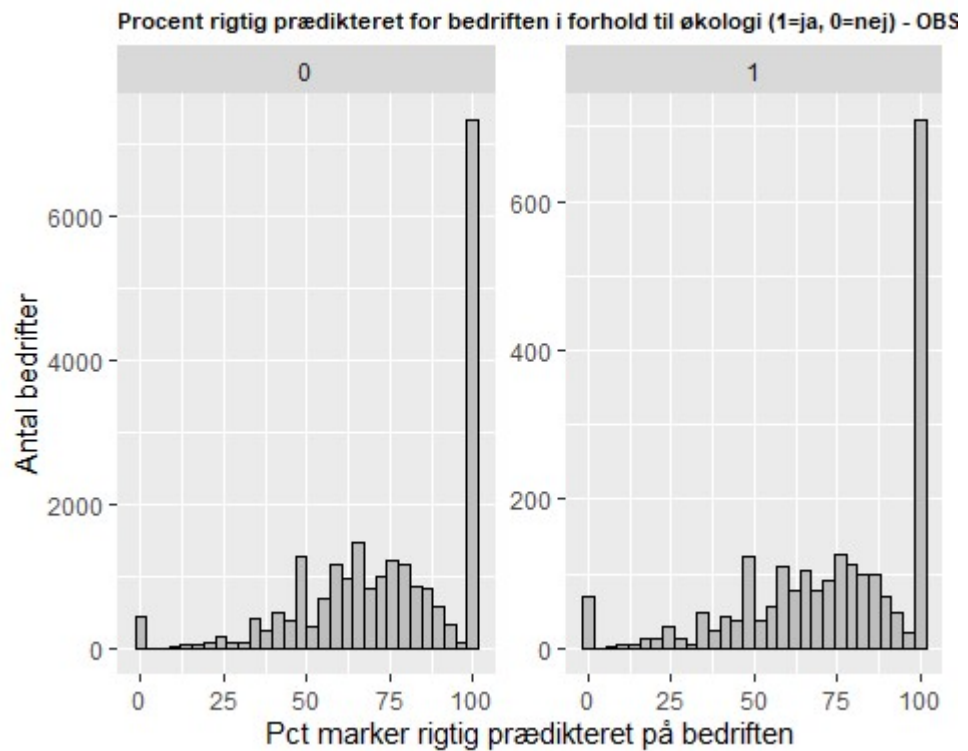
8.

Figur 9: Procent rigtig prædikeret for bedriften i forhold til bedriftstype



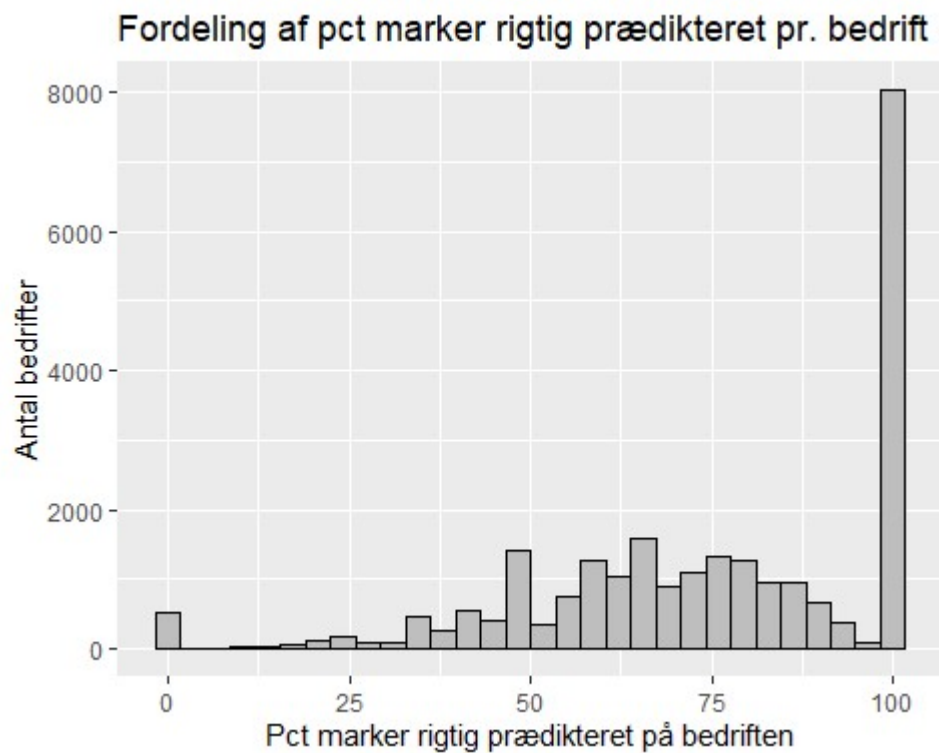
9.

Figur 10: Procent rigtig prædikeret for bedriften i forhold til region



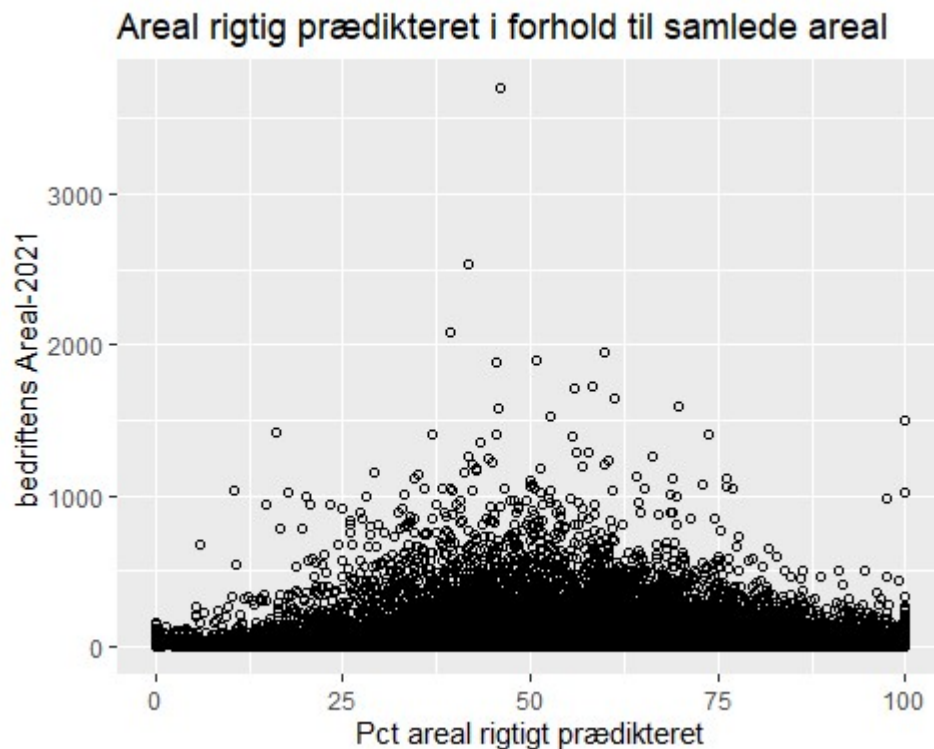
10.

Figur 11: Procent rigtig prædikeret for bedriften i forhold til økologi



11.

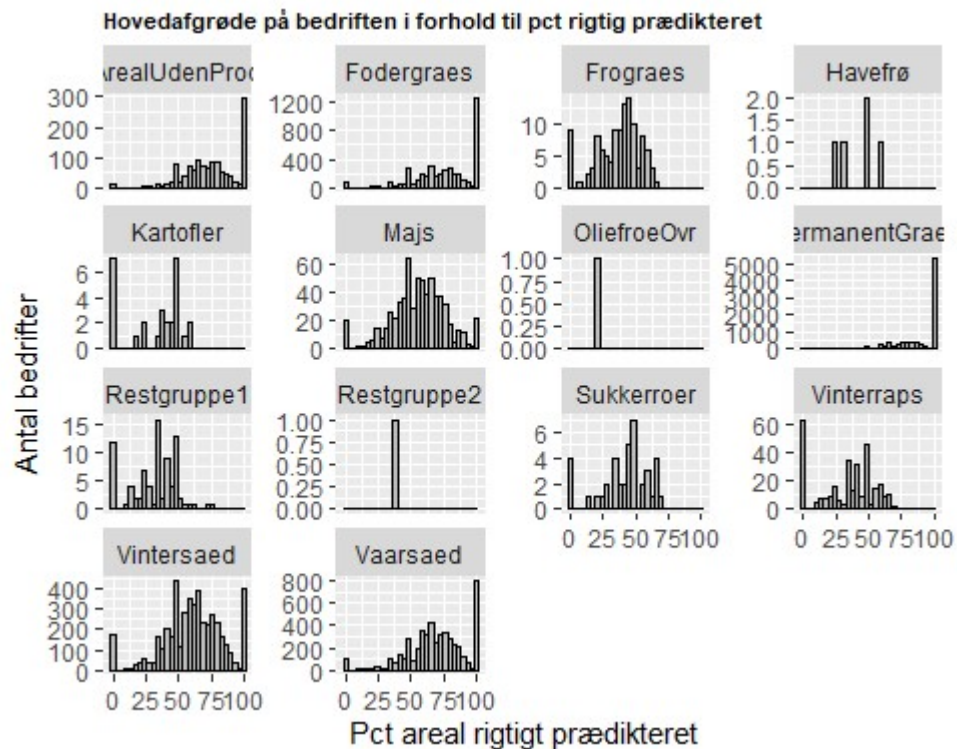
Figur 12: Fordeling af antal bedrifter i forhold til procent marker prædikeret rigtigt på bedriften



12.

Figur 13: Fordeling af procent rigtig prædikteret areal i forhold til bedriftens 2021-areal

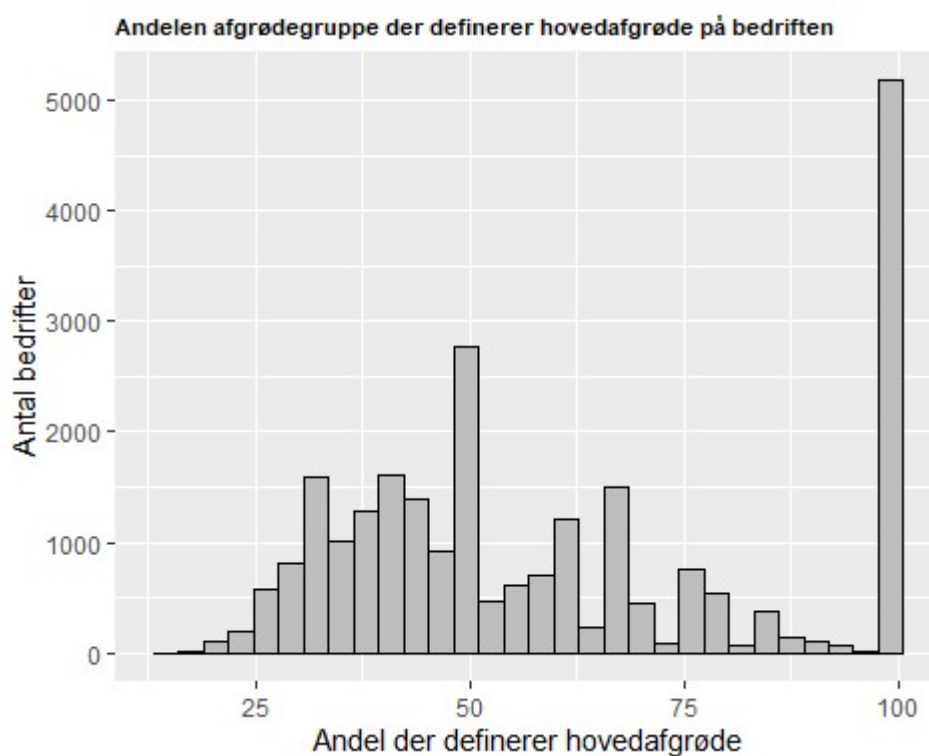
Der er set på hvilken hovedafgrøde der dominerer på bedriften. Bedriftens hovedafgrøde er defineret som den største andel af en afgrøde på bedriftens marker. Hvis der er to lige store andele, er der her tilfældigt valgt den ene. Nedenfor ses et plot af hvordan "rigtighedsprocenten" fordeler sig indenfor hovedafgrødegrupperne på bedriftniveau.



13.

Figur 14: Fordeling af procent rigtig prædikteret areal i forhold til bedriftens hovedafgrøde

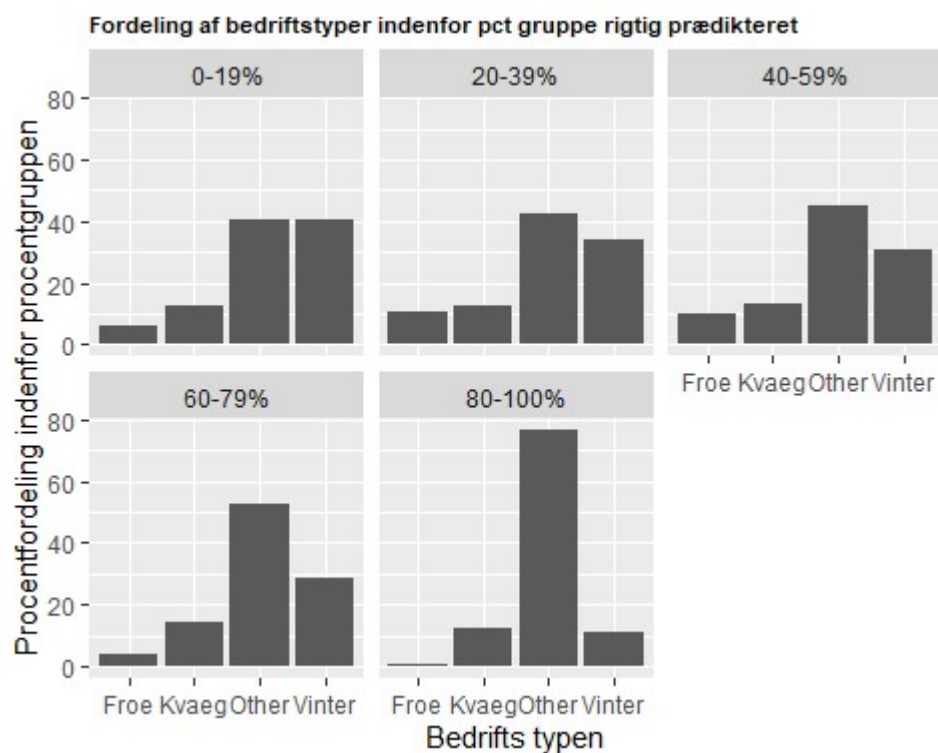
Hvor stor en andel der definerer en hovedafgrødegruppe på bedrifterne ses her nedenfor.



14.

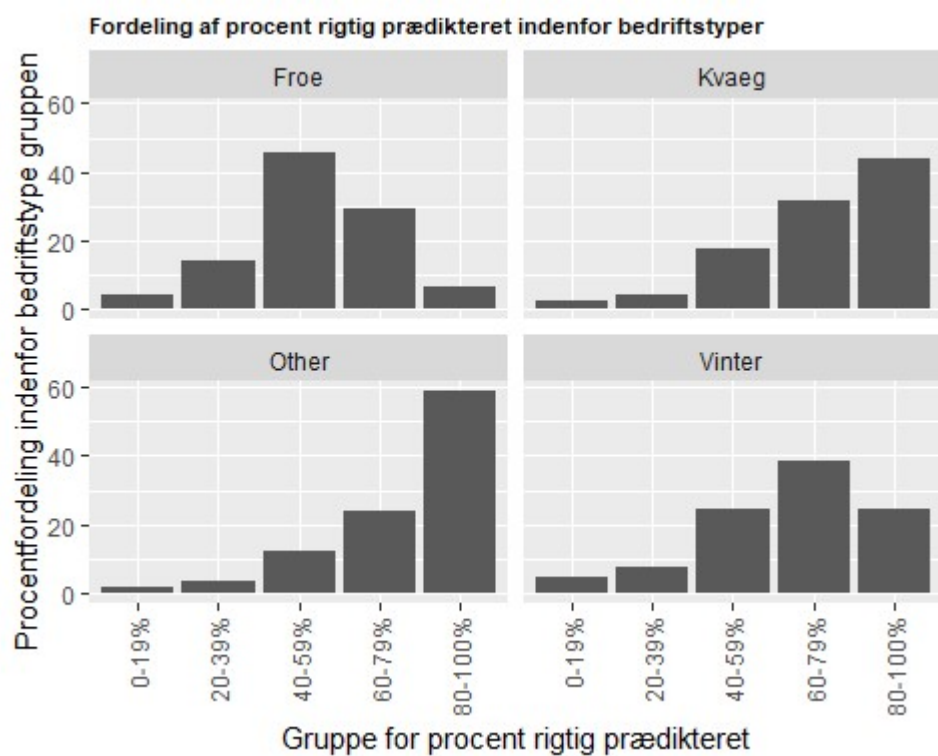
Figur 15: Andelen der definerer en bedrifts hovedafgrøde

Bedriftstyper er ligeledes kategoriseret ved om bedriftstypen er Frø, Kvæg, Vinter og andet. Nedenfor ses hvordan bedrifterne fordeler sig indenfor den forskellige grupper af "rigtighedsprocenten".



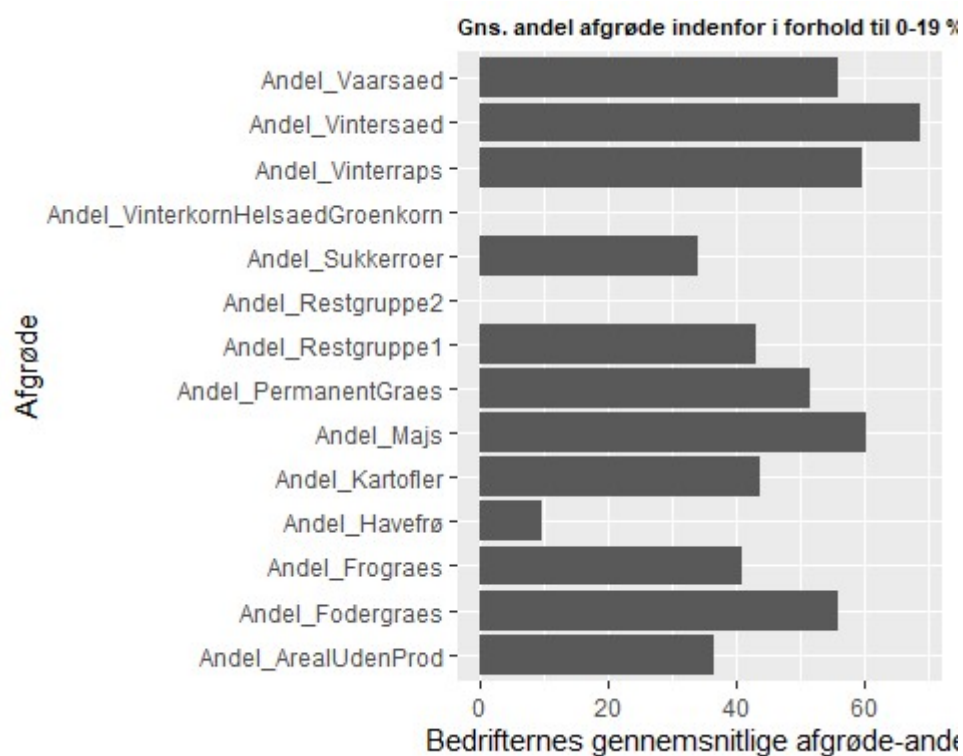
15.

Figur 16: Fordeling af bedriftstype i forhold til pct rigtig prædikeret



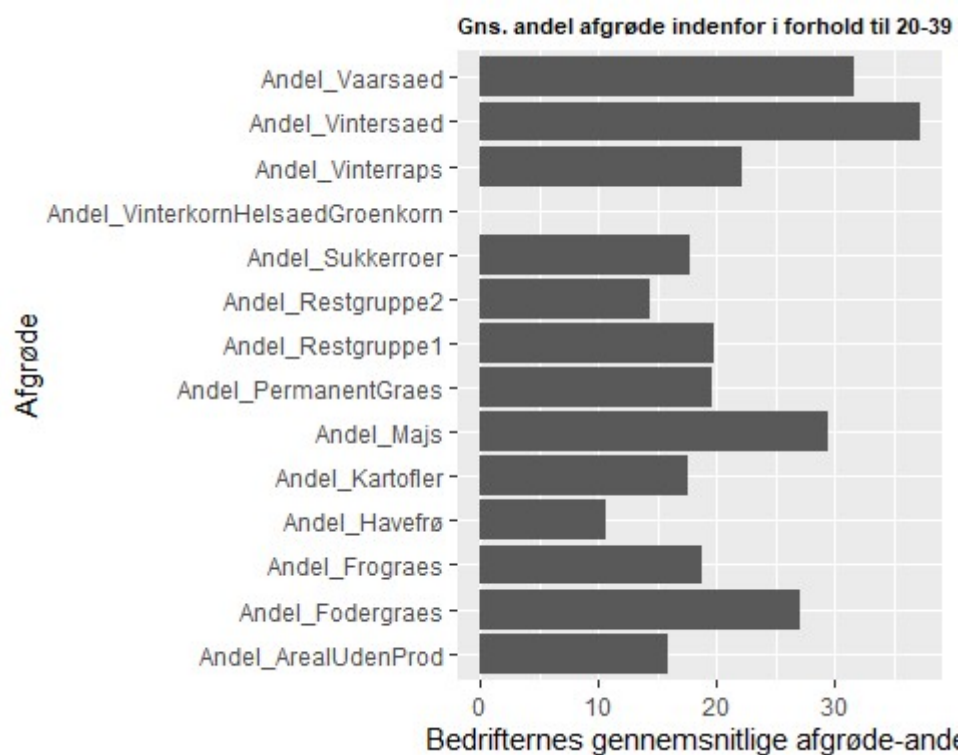
16.

Figur 17: Fordeling af pct rigtig prædikeret i forhold til bedriftstype



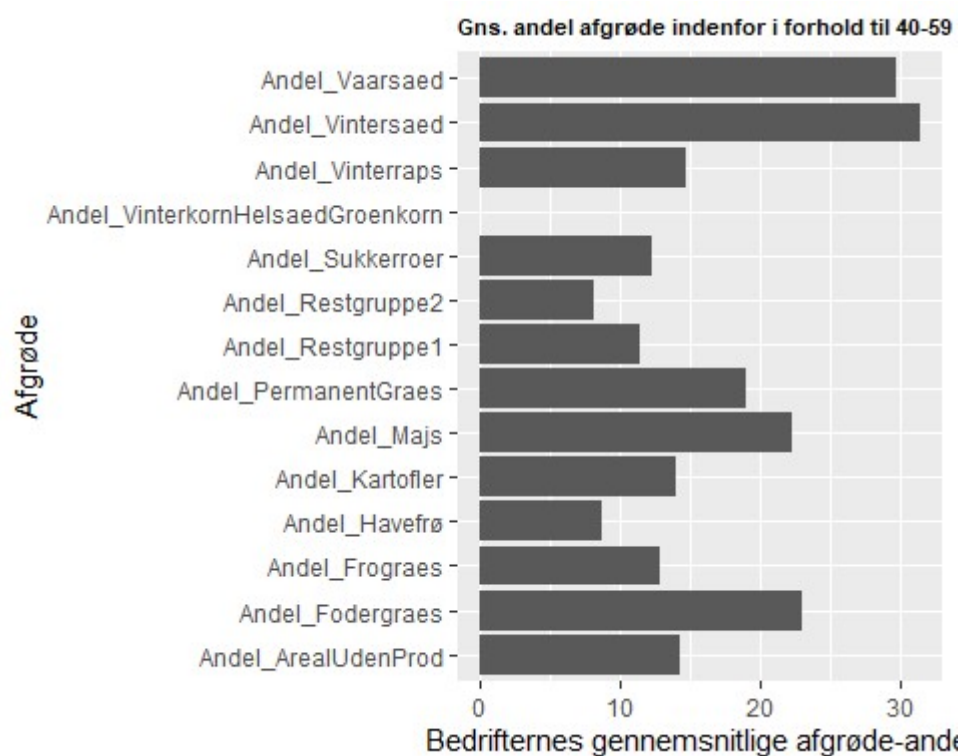
17.

Figur 18: Den gennemsnitlige andel af afgrødegruppe indenfor pct gruppe: 0-19%



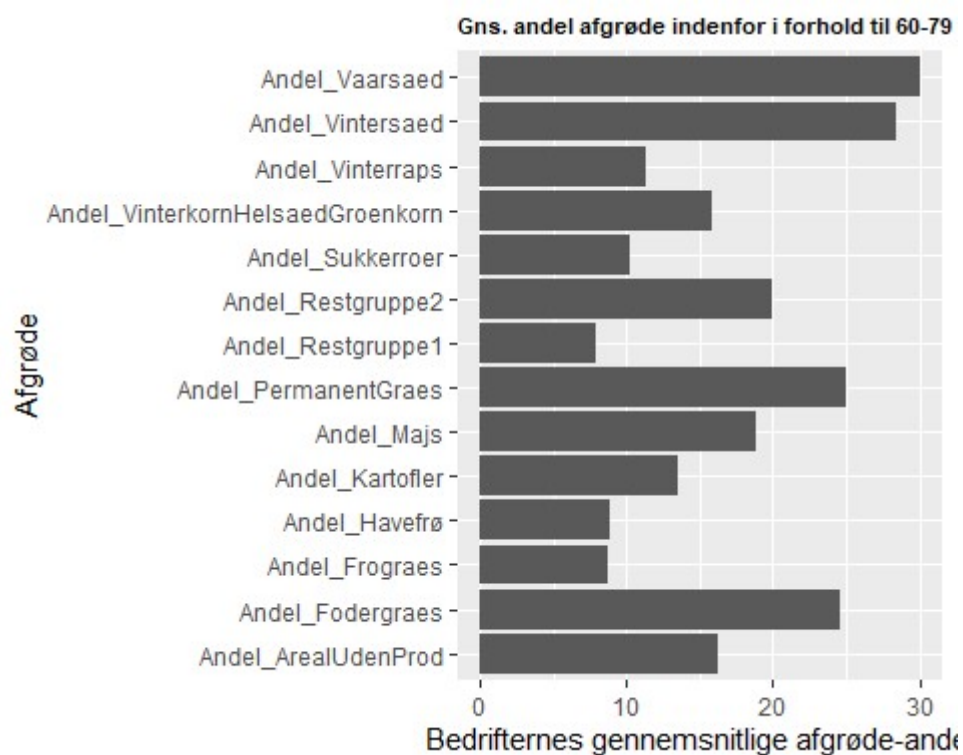
18.

Figur 19: Den gennemsnitlige andel af afgrødegruppe indenfor pct gruppe: 20-39%



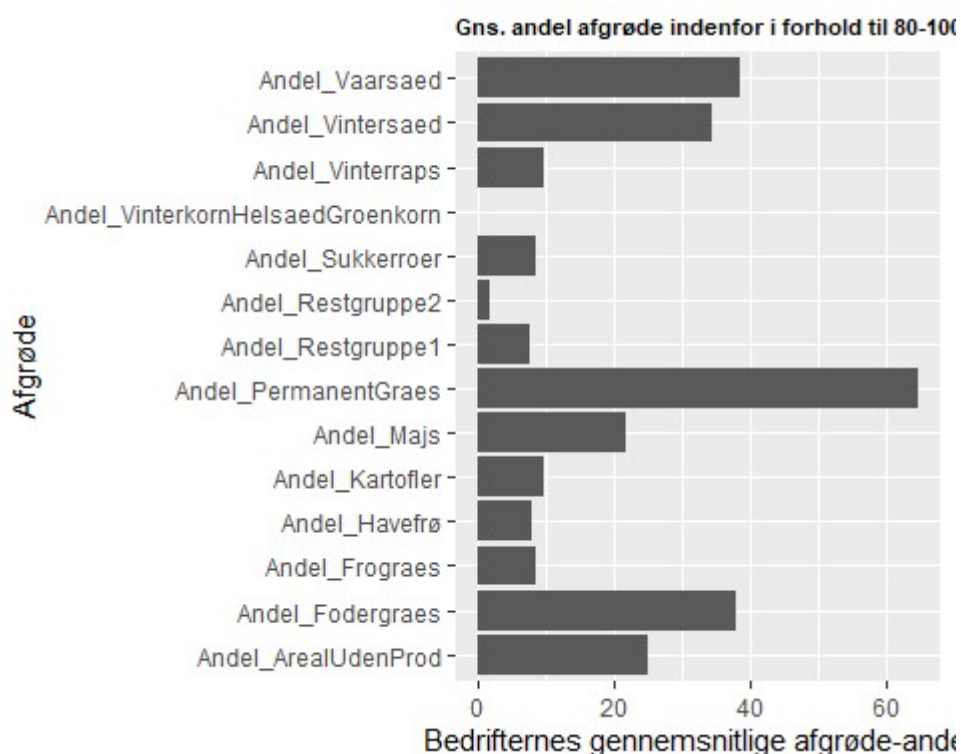
19.

Figur 20: Den gennemsnitlige andel af afgrødegruppe indenfor pct gruppe: 40-59%



20.

Figur 21: Den gennemsnitlige andel af afgrødegruppe indenfor pct gruppe: 60-79%



21.

Figur 22: Den gennemsnitlige andel af afgrødegruppe indenfor pct gruppe: 80-100%

Opsamling

Det er observeret at marker med flere observationer ikke prædikteres bedre. Ligeledes er der jo nogle afgrødegrupper som ikke prædikteres også selvom de ikke er små i antal af marker. Nogle af de afgrøder er dem hvor der typisk er en rotations cyklus. Det vil eksempelvis sige raps der har ca. 4 frie år i mellem. Dette fanges desværre ikke godt nok af denne type model.

Det er derfor ønskeligt at undersøge modeller der er endnu bedre til at fange sekvensmønstre, for dermed at blive bedre til at prædiktere afgrødegrupper der har frie år imellem. Her har der blandt andet været snakket om at undersøge en RNN model. I projektet her har det været brugt meget tid på dataklargøring og valget har været at starte med en mere simpel model. Der har været fokus på at koble modeloutputs til bedriftsoplysninger i denne analyse. Dette har givet mulighed for at definere nogle faglige vinkler og tiltag til datastruktur i modellen.

Det kunne fremover tænkes at en arealvægtning fremfor antal marker i forhold til ubalancen i afgrøder kunne give andre resultater og måske bedre inkludere afgrødegruppen raps.

Der har været fokus i starten på at marker der var 1 år ude (dvs. eks marker med bortforpagtning til eksempelvis kartofler) skulle tælle med i bedriftens historik. Grundet anden dataresourc prioritering er dette ikke blevet yderligere behandlet i programmeringen her. Dette medfører at kartoffelmarker med andet cvrnummer vil betragtes som et "ejerskifte" og dermed vil disse sekvenser med 1 års kartofler ikke være med. Efterfølgende ses i lyset af hvordan rapssekvenserne ikke er blevet rigtigt fanget af denne model, da må man forvente det vil være det samme for kartofler. Ved en yderligere iteration af modeludvælgelse da bør man prioritere denne problematik.

Der er fremsat et fremtidigt ønske om forsøgsvis at tage "arealer uden produktion" og "permanent græs" ud af modellen – fordi disse marker alligevel blot bliver liggende og ikke efterfølges af andet.

Der ønskes at vi i den sidste ende hellere vil prædiktere for få bedrifter rigtigt end at kunne ramme alle bedrifter. Derfor er der i analysen set på nogle potentielle karakteristika ved de

bedrifter hvor vi rammer godt, for måske kunne beskrive og definere denne gruppe. Det ser dog ud til at modellen er så god til at prædiktere de "permanente marker" hvor der ikke sker store skift, at de marker med eksempelvis permanent græs, fodergræs og areal uden produktion er de afgrødegrupper der fylder mest ved de bedrifter hvor vi prædikterer mellem 80-100 % rigtigt på markerne.

Håbet er at en anden model der bedre håndterer sekvensinput måske bedre fanger de andre afgrødegrupper, og vi dermed mere tydeligt ser nogle andre karakteristika ved de bedrifter.

Appendiks

En fuld beskrivelse af den indledende datamanagement del kan findes på U:drev *U:/Virkemiddelvælgeren/AfgPrognose/Dokumentation/Databeskrivelse*

Her beskrives rådatafilerne og hvorledes de er koblet sammen og data er defineret.

Som grundlag for analyserapporten her er brugt programmerne (*DataInput_vers3_2.R*, *Data-management_vers3_2.R*, *Models_vers3.R*, *Models_Test3.R*).

I dette notat er kort skitseret hvilke datavalg der er taget. Vi ser kun på bedrifter med kendt cvr-nummer i 2022. (for at kunne validere på data vælger vi at skulle kunne følge bedriften frem til 2022). Vi kræver minimum 2 observationer til prædiktionen. Der er lavet særlig filtreringer for "AreallkkeRetur", "PyntegrøntJuletræer", "EnergiafgrødeLavskov", "Kartofler" samt "Gartneriafgrøder".

Efter de filtreringer og datavalg der er, da er vi endt på 366051 marker fordelt på 24793 cvr-numre i træningsdatasættet.

Bedriftstyperne defineres som "Froe" hvis arealandelen af frø ("Frograes", "Spinat", "Havefrø") er mere end 15 %, "Kvaeg" hvis arealandelsen af kvaeg ("Fodergraes", "Majs") er mere end 50 %, "Vinter" hvis arealandelsen af vinter ("Vinter") er mere end 50 % og ellers "Other".

Nedenfor ses afgrødegrupperne og hvilke enkeltafgrøder disse indeholder.

Tabel 14: Afgrødegrupperne og deres enkeltafgrøder

AfgKode	AfgNavn	GrpKode	GrpNavn
1	Vårbyg	Grp_1	Vaarsaed
3	Vårhavre	Grp_1	Vaarsaed
4	Blanding af vårsæde arter	Grp_1	Vaarsaed
2	Vårhvede	Grp_1	Vaarsaed
7	Korn + bælgssæd under 50% bælgssæd	Grp_1	Vaarsaed
6	Vårhvede, brødhvede	Grp_1	Vaarsaed
58	Sorghum	Grp_1	Vaarsaed
56	Vårtriticale	Grp_1	Vaarsaed
55	Vårrug	Grp_1	Vaarsaed
8	Vårspelt	Grp_1	Vaarsaed
230	Blanding af vårkorn, grønkorn	Grp_1	Vaarsaed
214	Korn og bælgssæd, helsæd, under 50% bælgssæd	Grp_1	Vaarsaed
210	Vårbyg, helsæd	Grp_1	Vaarsaed
215	Ærtehelssæd	Grp_1	Vaarsaed
211	Vårhvede, helsæd	Grp_1	Vaarsaed
212	Vårhavre, helsæd	Grp_1	Vaarsaed
213	Blandkorn, vårsæt, helsæd	Grp_1	Vaarsaed
234	Korn og bælgssæd, grønkorn, under 50% bælgssæd	Grp_1	Vaarsaed
701	Grønkorn af vårbyg	Grp_1	Vaarsaed
702	Grønkorn af vårhvede	Grp_1	Vaarsaed
704	Grønkorn af vårrug	Grp_1	Vaarsaed
703	Grønkorn af vårhavre	Grp_1	Vaarsaed
705	Grønkorn af vårtriticale	Grp_1	Vaarsaed
217	Korn og bælgssæd, helsæd (over 50 % bælgssæd)	Grp_1	Vaarsaed
18	Korn og bælgssæd over 50% bælgssæd	Grp_1	Vaarsaed
101	Rajgræsfrø, alm.	Grp_101	Frograes
114	Rapgræsfrø, alm.	Grp_101	Frograes
120	Kløverfrø	Grp_101	Frograes
103	Rajgræsfrø, ital.	Grp_101	Frograes
126	Blanding af markfrø til udsæd	Grp_101	Frograes
116	Rajgræs, hybrid	Grp_101	Frograes
105	Timothefrø	Grp_101	Frograes
121	Bælgplanter, frø	Grp_101	Frograes
123	Valmuefrø	Grp_101	Frograes
108	Rødsvingelfrø	Grp_101	Frograes
113	Engrapsgræsfrø (plænetype)	Grp_101	Frograes

AfgKode	AfgNavn	GrpKode	GrpNavn
107	Engsvingelfrø	Grp_101	Frograes
112	Engrapgræsfrø (marktype)	Grp_101	Frograes
111	Svingelfrø, strand-	Grp_101	Frograes
106	Hundegræsfrø	Grp_101	Frograes
109	Rajsvingelfrø	Grp_101	Frograes
110	Svingelfrø, bakke- (tidl. Stivbladet)	Grp_101	Frograes
115	Hvenerfrø, alm. og krybende	Grp_101	Frograes
125	Bederøfrø	Grp_101	Frograes
102	Rajgræsfrø, alm. 1. år, efterårsudlagt	Grp_101	Frograes
104	Rajgræsfrø, ital. 1. år efterårsudlagt	Grp_101	Frograes
117	Rajgræs, efterårsudl. hybrid	Grp_101	Frograes
122	Kommenfrø	Grp_101	Frograes
118	Rajsvingelfrø, efterårsudlagt	Grp_101	Frograes
10	Vinterbyg	Grp_11	Vintersaed
11	Vinterhvede	Grp_11	Vintersaed
13	Vinterhvede, brødhvede	Grp_11	Vintersaed
14	Vinterrug	Grp_11	Vintersaed
15	Vinterhybridrug	Grp_11	Vintersaed
16	Vintertriticale	Grp_11	Vintersaed
235	Blanding af vinterkorn, grønkorn	Grp_11	Vintersaed
221	Vinterhvede, helsæd	Grp_11	Vintersaed
220	Vinterbyg, helsæd	Grp_11	Vintersaed
222	Vinterrug, helsæd	Grp_11	Vintersaed
224	Blandkorn, efterårssæet helsæd	Grp_11	Vintersaed
223	Vintertriticale, helsæd	Grp_11	Vintersaed
709	Grønkorn af vinterrug	Grp_11	Vintersaed
707	Grønkorn af vinterhvede	Grp_11	Vintersaed
706	Grønkorn af vinterbyg	Grp_11	Vintersaed
17	Blanding af efterårssæede arter	Grp_11	Vintersaed
57	Vinterhavre	Grp_11	Vintersaed
9	Vinterspelt	Grp_11	Vintersaed
151	Kartofler, stivelses-	Grp_151	Kartofler
152	Kartofler, spise- (pakkeri, vejsalg)	Grp_151	Kartofler
150	Kartofler, lægge- (egen opformering)	Grp_151	Kartofler
149	Kartofler, lægge- (certificerede)	Grp_151	Kartofler
156	Kartofler, friteret/chips/pommes frites	Grp_151	Kartofler

AfgKode	AfgNavn	GrpKode	GrpNavn
155	Kartofler, pulver/granules-	Grp_151	Kartofler
154	Kartofler, spise- (proces, skrællede kogte)	Grp_151	Kartofler
153	Kartofler, andre	Grp_151	Kartofler
160	Sukkerroer til fabrik	Grp_160	Sukkerroer
280	Fodersukkerroer	Grp_160	Sukkerroer
282	Fodermarkål	Grp_160	Sukkerroer
281	Kålroer	Grp_160	Sukkerroer
161	Cikorierødder	Grp_160	Sukkerroer
162	Blanding, andre industriafrgr.	Grp_160	Sukkerroer
180	Gul sennep	Grp_21	OliefroeOvr
181	Anden oliefrøart	Grp_21	OliefroeOvr
182	Blanding af oliearter	Grp_21	OliefroeOvr
5	Majs til modenhed	Grp_216	Majs
216	Silomajs	Grp_216	Majs
22	Vinterraps	Grp_22	Vinterraps
252	Permanent græs, normalt udbytte	Grp_252	PermanentGraes
251	Permanent græs, lavt udbytte	Grp_252	PermanentGraes
276	Permanent græs og kløvergræs uden norm, under 50 % kløver	Grp_252	PermanentGraes
250	Permanent græs, meget lavt udbytte	Grp_252	PermanentGraes
259	Permanent græs, fabrik, over 6 tons	Grp_252	PermanentGraes
286	Permanent græs og kløvergræs uden norm, over 50 % kløver	Grp_252	PermanentGraes
287	Græs til udegrise, permanent	Grp_252	PermanentGraes
254	Miljøgræs MVJ-tilsagn (0 N), permanent	Grp_252	PermanentGraes
253	Miljøgræs MVJ-tilsagn (80 N), omdrift	Grp_252	PermanentGraes
247	Miljøgræs MVJ-tilsagn (0 N), omdrift	Grp_252	PermanentGraes
255	Permanent græs, under 50% kløver/lucerne	Grp_252	PermanentGraes
257	Permanent græs, uden kløver	Grp_252	PermanentGraes
256	Permanent kløvergræs, over 50% kløver/lucerne	Grp_252	PermanentGraes
279	Permanent kløvergræs til fabrik	Grp_252	PermanentGraes
278	Permanent lucerne og lucernegræs over 50% lucerne	Grp_252	PermanentGraes
272	Permanent græs til fabrik	Grp_252	PermanentGraes
274	Permanent lucernegræs over 25% græs, til fabrik	Grp_252	PermanentGraes
270	Græs til udegrise, omdrift	Grp_252	PermanentGraes
269	Græs, rullegræs	Grp_252	PermanentGraes
248	Permanent græs ved vandboring	Grp_252	PermanentGraes
249	Udnyttet græs ved vandboring	Grp_252	PermanentGraes

AfgKode	AfgNavn	GrpKode	GrpNavn
258	Permanent græs, ø-støtte	Grp_252	PermanentGraes
350	Miljøgræs med N-kvot	Grp_254	GraesMedPleje
260	Græs med kløver/lucerne, under 50 % bælgpl. (omdrift)	Grp_260	Fodergraes
268	Græs under 50% kløver/lucerne, lavt udbytte (omdrift)	Grp_260	Fodergraes
264	Græs og kløvergræs uden norm, under 50 % kløver (omdrift)	Grp_260	Fodergraes
267	Græs under 50% kløver/lucerne, meget lavt udbytte (omdrift)	Grp_260	Fodergraes
263	Græs uden kløvergræs (omdrift)	Grp_260	Fodergraes
266	Græs under 50% kløver/lucerne, ekstremt lavt udbytte (omdrift)	Grp_260	Fodergraes
261	Kløvergræs, over 50% kløver (omdrift)	Grp_260	Fodergraes
172	Lucernegræs, over 25% græs til slæt inkl. eget foder	Grp_260	Fodergraes
171	Lucerne, slæt	Grp_260	Fodergraes
174	Kløvergræs til fabrik	Grp_260	Fodergraes
170	Græs til fabrik (omdrift)	Grp_260	Fodergraes
173	Kløver til slæt	Grp_260	Fodergraes
262	Lucernegræs, over 50% lucerne (omdrift)	Grp_260	Fodergraes
284	Græs med vikke og andre bælgplanter, under 50 % bælgpl.	Grp_260	Fodergraes
285	Græs og kløvergræs uden norm, over 50 % kløver (omdrift)	Grp_260	Fodergraes
236	Græs med kløver/lucerne, under 50 % bælgpl. (omdrift) efterårsudlagt i vin	Grp_260	Fodergraes
237	Græs med kløver/lucerne, over 50 % bælgpl. (omdrift) efterårsudlagt i vint	Grp_260	Fodergraes
265	Græs, slæt før vårsået afgrøde	Grp_260	Fodergraes
273	Lucerne til fabrik	Grp_260	Fodergraes
277	Kløver til fabrik	Grp_260	Fodergraes
275	Permanent rullegræs	Grp_270	SærligtGraes
283	Fodergulerødder	Grp_280	Foderroer
310	Brak, sommerslåning	Grp_308	ArealUdenProd
271	Rekreative formål	Grp_308	ArealUdenProd
319	MFO-brak, Udtagning, ej landbrugsareal	Grp_308	ArealUdenProd
579	Tagetes, sygdomssanerende plante	Grp_308	ArealUdenProd
308	MFO-brak, sommerslåning	Grp_308	ArealUdenProd
306	Græs i omdrift, uden udbetaling af økologi-tilskud	Grp_308	ArealUdenProd
305	Permanent græs, uden udbetaling af økologi-tilskud	Grp_308	ArealUdenProd
325	MFO-Blomsterbrak	Grp_308	ArealUdenProd
324	Blomsterbrak	Grp_308	ArealUdenProd
339	MFO-brak, forårsslåning	Grp_308	ArealUdenProd
327	MFO-brømme, sommerslåning	Grp_308	ArealUdenProd
334	MFO-brømme, forårsslåning	Grp_308	ArealUdenProd

AfgKode	AfgNavn	GrpKode	GrpNavn
338	Brak, forårsslåning	Grp_308	ArealUdenProd
343	MFO-bestøverbrak	Grp_308	ArealUdenProd
328	MFO-bræmme med blomsterblanding	Grp_308	ArealUdenProd
336	MFO-bræmme, permanent græs, sommerslåning	Grp_308	ArealUdenProd
342	Bestøverbrak	Grp_308	ArealUdenProd
329	MFO-bræmme, miljøtilsagn	Grp_308	ArealUdenProd
335	MFO-bræmme, permanent græs, forårsslåning	Grp_308	ArealUdenProd
345	Brak langs vandløb og søer, sommerslåning (alternativ til efterafgrøder)	Grp_308	ArealUdenProd
344	Brak langs vandløb og søer, forårsslåning (alternativ til efterafgrøder)	Grp_308	ArealUdenProd
337	MFO-bræmme, permanent græs, miljøtilsagn	Grp_308	ArealUdenProd
309	Udyrket areal ved vandboring	Grp_308	ArealUdenProd
315	Miljøgræs med udtagning	Grp_308	ArealUdenProd
320	Braklagte randzoner	Grp_308	ArealUdenProd
323	MFO-udyrket areal ved vandboring	Grp_308	ArealUdenProd
326	Ej landbrug, MSO, omlagt fra permanent græs	Grp_308	ArealUdenProd
340	Randzoneordningen	Grp_308	ArealUdenProd
341	Randzoneordning med udtagning	Grp_308	ArealUdenProd
360	Vildtafgrøder	Grp_308	ArealUdenProd
486	Hønsegård uden plantedække	Grp_308	ArealUdenProd
487	Skovlandbrug	Grp_308	ArealUdenProd
488	Hønsegård, permanent græs	Grp_308	ArealUdenProd
888	Nye tilsagn uden råderet v. ansøgningsfristen	Grp_308	ArealUdenProd
900	Øvrige afgrøder	Grp_308	ArealUdenProd
901	Naturarealer efter driftsloven	Grp_308	ArealUdenProd
902	Naturlignende arealer	Grp_308	ArealUdenProd
903	Lysåbne arealer i skov	Grp_308	ArealUdenProd
904	Åbne arealer i skov	Grp_308	ArealUdenProd
905	Anden anvendelse på tilsagnsarealer	Grp_308	ArealUdenProd
906	Afmeldte arealer	Grp_308	ArealUdenProd
907	Naturarealer, økologisk jordbrug	Grp_308	ArealUdenProd
908	Naturarealer, ansøgning om miljøtilsagn	Grp_308	ArealUdenProd
910	Ikke anmeldt mark	Grp_308	ArealUdenProd
920	Økologisk sommerbrak	Grp_308	ArealUdenProd
921	Bar jord	Grp_308	ArealUdenProd
124	Spinatfrø	Grp_31	Restgruppe1
21	Vårrops	Grp_31	Restgruppe1

AfgKode	AfgNavn	GrpKode	GrpNavn
24	Solsikke	Grp_31	Restgruppe1
23	Rybs	Grp_31	Restgruppe1
32	Sødlupin	Grp_31	Restgruppe1
36	Bælgsæd, andre typer til modenhed blanding	Grp_31	Restgruppe1
424	Ærter, konsum	Grp_31	Restgruppe1
31	Hestebønner	Grp_31	Restgruppe1
30	Ærter	Grp_31	Restgruppe1
25	Sojabønner	Grp_31	Restgruppe1
35	Bælgsæd, flerårig blanding	Grp_31	Restgruppe1
54	Bælgsæd blanding	Grp_31	Restgruppe1
311	Skovrejsning på tidl. landbrugsjord 1	Grp_318	AreallkkeRetur
318	MVJ ej udtagning, ej landbrugsareal	Grp_318	AreallkkeRetur
312	20-årig udtagning	Grp_318	AreallkkeRetur
313	20-årig udtagning af agerjord med frivillig skovrejsning	Grp_318	AreallkkeRetur
317	Vådområder med udtagning	Grp_318	AreallkkeRetur
321	Miljøtiltag, ej landbrugsarealer	Grp_318	AreallkkeRetur
361	Ikke støtteberettiget landbrugsareal	Grp_318	AreallkkeRetur
316	20-årig Udtagning med fastholdelse, ej landbrugsareal	Grp_318	AreallkkeRetur
322	Minivådområder, projektilsagn	Grp_318	AreallkkeRetur
314	20-årig udtagning med tilsagn om skovrejsning	Grp_318	AreallkkeRetur
575	Skovrejsning (privat) ? kulstofbinding og grundvandsbeskyttelse	Grp_318	AreallkkeRetur
576	Skovrejsning (statslig) - forbedring af vandmiljø og grundvandsbeskyttelse	Grp_318	AreallkkeRetur
577	Skov med biodiversitetsformål	Grp_318	AreallkkeRetur
578	Skovrejsning (privat) - forbedring af vandmiljø og grundvandsbeskyttelse	Grp_318	AreallkkeRetur
580	Anden skovdrift	Grp_318	AreallkkeRetur
581	Skovdrift med fjernelse af ved	Grp_318	AreallkkeRetur
585	Skovrejsning i projektområde, som ikke er omfattet af tilsagn	Grp_318	AreallkkeRetur
586	Offentlig skovrejsning	Grp_318	AreallkkeRetur
587	Skovrejsning på tidl. landbrugsjord 3	Grp_318	AreallkkeRetur
588	Statslig skovrejsning	Grp_318	AreallkkeRetur
589	Bæredygtig skovdrift	Grp_318	AreallkkeRetur
590	Bæredygtig skovdrift i Natura 2000-område	Grp_318	AreallkkeRetur
50	Anden bredbladet afgrøde	Grp_42	Restgruppe2
40	Oliehør	Grp_42	Restgruppe2
42	Hamp	Grp_42	Restgruppe2
52	Quinoa	Grp_42	Restgruppe2

AfgKode	AfgNavn	GrpKode	GrpNavn
51	Blanding bredbladet afgrøde, frø/kerne	Grp_42	Restgruppe2
53	Boghvede	Grp_42	Restgruppe2
41	Spindhør	Grp_42	Restgruppe2
411	Løg	Grp_420	GartneriAfg
407	Gulerod	Grp_420	GartneriAfg
420	Salat (friland)	Grp_420	GartneriAfg
415	Porre	Grp_420	GartneriAfg
408	Hvidkål	Grp_420	GartneriAfg
401	Asparges	Grp_420	GartneriAfg
412	Pastinak	Grp_420	GartneriAfg
406	Grønkål	Grp_420	GartneriAfg
403	Blomkål	Grp_420	GartneriAfg
423	Suktermajs	Grp_420	GartneriAfg
422	Spinat	Grp_420	GartneriAfg
430	Bladpersille	Grp_420	GartneriAfg
429	Jordskokker, konsum	Grp_420	GartneriAfg
402	Bladselleri	Grp_420	GartneriAfg
400	Asieagurker	Grp_420	GartneriAfg
417	Rødbede	Grp_420	GartneriAfg
418	Rødkål	Grp_420	GartneriAfg
421	Savoykål, spidskål	Grp_420	GartneriAfg
410	Knoldselleri	Grp_420	GartneriAfg
404	Broccoli	Grp_420	GartneriAfg
413	Rodpersille	Grp_420	GartneriAfg
405	Courgette, squash	Grp_420	GartneriAfg
431	Purløg	Grp_420	GartneriAfg
432	Krydderurter (undtagen persille og purløg)	Grp_420	GartneriAfg
416	Rosenkål	Grp_420	GartneriAfg
409	Kinakål	Grp_420	GartneriAfg
434	Grøntsager, andre (friland)	Grp_420	GartneriAfg
433	Krydderurter, andre	Grp_420	GartneriAfg
440	Solhat	Grp_420	GartneriAfg
448	Medicinpl., en- og toårige	Grp_420	GartneriAfg
449	Medicinpl., stauder	Grp_420	GartneriAfg
450	Grøntsager, blandinger	Grp_420	GartneriAfg
0		Grp_44	INTET NAVN

AfgKode	AfgNavn	GrpKode	GrpNavn
44	INTEt NAVN	Grp_44	INTEt NAVN
200	Fragået mark	Grp_44	INTEt NAVN
201	Sammenlagt mark	Grp_44	INTEt NAVN
202	fragået mark, prod.skifte	Grp_44	INTEt NAVN
218	INTEt NAVN	Grp_44	INTEt NAVN
300	INTEt NAVN	Grp_44	INTEt NAVN
800	INTEt NAVN	Grp_44	INTEt NAVN
801	INTEt NAVN	Grp_44	INTEt NAVN
961	INTEt NAVN	Grp_44	INTEt NAVN
962	INTEt NAVN	Grp_44	INTEt NAVN
965	INTEt NAVN	Grp_44	INTEt NAVN
998	Ukendt afgrøde	Grp_44	INTEt NAVN
999	INTEt NAVN	Grp_44	INTEt NAVN
1000	Ingen afgrødetekst	Grp_44	INTEt NAVN
489	Havtorn	Grp_489	Planteskole
490	Hassel, træ (Corylus avellana)	Grp_489	Planteskole
491	Storfrugtet tranebær	Grp_489	Planteskole
493	Surbær	Grp_489	Planteskole
494	Japan kvæde	Grp_489	Planteskole
495	Morbær	Grp_489	Planteskole
496	Medicinpl., vedplanter	Grp_489	Planteskole
497	Planteskolekulturer, vedplanter, til videresalg	Grp_489	Planteskole
498	Containerplads 4, vedplanter	Grp_489	Planteskole
499	Lukket system	Grp_489	Planteskole
500	Buske og træer	Grp_489	Planteskole
501	Stauder	Grp_489	Planteskole
502	Blomsterløg	Grp_489	Planteskole
503	En- og to-årige planter	Grp_489	Planteskole
504	Solbær, stiklingeopformering	Grp_489	Planteskole
505	Ribs, stiklingeopformering	Grp_489	Planteskole
506	Stikkelsbær, stiklingeopformering	Grp_489	Planteskole
507	Hindbær, stiklingeopformering	Grp_489	Planteskole
508	Andre af slægten Vaccinium	Grp_489	Planteskole
509	Trækvæde	Grp_489	Planteskole
540	Tomater	Grp_489	Planteskole
541	Agurker	Grp_489	Planteskole
542	Salat (drivhus)	Grp_489	Planteskole

AfgKode	AfgNavn	GrpKode	GrpNavn
543	Grøntsager, andre (drivhus)	Grp_489	Planteskole
544	Snitblomster og snitgrønt	Grp_489	Planteskole
545	Potteplanter	Grp_489	Planteskole
547	Planteskolekulturer, stauder	Grp_489	Planteskole
548	Småplanter, en-årige	Grp_489	Planteskole
549	Lukket system 1, en-årige	Grp_489	Planteskole
550	Lukket system 2, stauder	Grp_489	Planteskole
562	Containerplads 3, stauder	Grp_489	Planteskole
563	Svampe, champignon	Grp_489	Planteskole
564	Containerplads	Grp_489	Planteskole
510	Melon	Grp_510	FrugtBær
511	Græskar	Grp_510	FrugtBær
512	Rabarber	Grp_510	FrugtBær
513	Jordbær	Grp_510	FrugtBær
514	Solbær	Grp_510	FrugtBær
515	Ribs	Grp_510	FrugtBær
516	Stikkelsbær	Grp_510	FrugtBær
517	Brombær	Grp_510	FrugtBær
518	Hindbær	Grp_510	FrugtBær
519	Blåbær	Grp_510	FrugtBær
520	Surkirsebær uden undervækst af græs	Grp_510	FrugtBær
521	Surkirsebær med undervækst af græs	Grp_510	FrugtBær
522	Blomme uden undervækst af græs	Grp_510	FrugtBær
523	Blomme med undervækst af græs	Grp_510	FrugtBær
524	Sødkirsebær uden undervækst af græs	Grp_510	FrugtBær
525	Sødkirsebær med undervækst af græs	Grp_510	FrugtBær
526	Hyd	Grp_510	FrugtBær
527	Hassel (Corylus maxima)	Grp_510	FrugtBær
528	Æbler	Grp_510	FrugtBær
529	Pærer	Grp_510	FrugtBær
530	Vindrue	Grp_510	FrugtBær
531	Anden træfrugt	Grp_510	FrugtBær
532	Anden buskfrugt	Grp_510	FrugtBær
533	Rønnebær	Grp_510	FrugtBær
534	Hyben	Grp_510	FrugtBær
535	Bærmispel	Grp_510	FrugtBær

AfgKode	AfgNavn	GrpKode	GrpNavn
536	Spisedruer	Grp_510	FrugtBær
537	Valnød (almindelig)	Grp_510	FrugtBær
538	Kastanje (ægte)	Grp_510	FrugtBær
539	Blandet frugt	Grp_510	FrugtBær
551	Moskusgræskar	Grp_510	FrugtBær
552	Mandelgræskar	Grp_510	FrugtBær
553	Centnergræskar	Grp_510	FrugtBær
560	Containerplads 1, frugtbuske	Grp_510	FrugtBær
561	Containerplads 2, en-årige	Grp_510	FrugtBær
570	Humle	Grp_510	FrugtBær
591	Lavskov	Grp_519	EnergiafgrødeLavskov
592	Pil	Grp_519	EnergiafgrødeLavskov
593	Poppel (0-100 andre træer pr. ha)	Grp_519	EnergiafgrødeLavskov
594	EI	Grp_519	EnergiafgrødeLavskov
596	Elefantgræs	Grp_519	EnergiafgrødeLavskov
597	Rørgræs	Grp_519	EnergiafgrødeLavskov
599	Poppel (100-400 andre træer pr. ha)	Grp_519	EnergiafgrødeLavskov
602	MFO - Pil	Grp_519	EnergiafgrødeLavskov
603	MFO - Poppel (0-100 andre træer pr. ha)	Grp_519	EnergiafgrødeLavskov
604	MFO - EI	Grp_519	EnergiafgrødeLavskov
605	MFO - Lavskov	Grp_519	EnergiafgrødeLavskov
606	MFO - Poppel (100-400 andre træer pr. ha)	Grp_519	EnergiafgrødeLavskov
582	Pyntegrønt, økologisk jordbrug	Grp_582	PyntegrøntJuletræer
583	Juletræer og pyntegrønt	Grp_582	PyntegrøntJuletræer
584	Juletræer og pyntegrønt i fredskov	Grp_582	PyntegrøntJuletræer
650	Chrysanthemum Garland, frø	Grp_650	Havefrø
651	Dildfrø	Grp_650	Havefrø
652	Kinesisk kålfrø	Grp_650	Havefrø
653	Karsefrø	Grp_650	Havefrø
654	Rucolafrø	Grp_650	Havefrø
655	Radisefrø (inklusive olieræddikefrø)	Grp_650	Havefrø
656	Bladbedefrø, rødbedefrø	Grp_650	Havefrø
657	Grønkålfrø	Grp_650	Havefrø
658	Gulerodsfrø	Grp_650	Havefrø
659	Kålfrø (hvid- og rødkål)	Grp_650	Havefrø
660	Persillefrø	Grp_650	Havefrø

AfgKode	AfgNavn	GrpKode	GrpNavn
661	Kørvelfrø	Grp_650	Havefrø
662	Majroefrø	Grp_650	Havefrø
663	Pastinakfrø	Grp_650	Havefrø
664	Skorzonerrod/skorzonerrodfrø	Grp_650	Havefrø
665	Havrerodfrø	Grp_650	Havefrø
666	Purløgsfrø	Grp_650	Havefrø
667	Timianfrø	Grp_650	Havefrø
668	Blomsterfrø	Grp_650	Havefrø
669	Andet havefrø	Grp_650	Havefrø
708	Grønkorn af vinterhavre	Grp_707	VinterkornHelsaedGroenkorn
710	Grønkorn af hybridrug	Grp_707	VinterkornHelsaedGroenkorn
711	Grønkorn af vintertriticale	Grp_707	VinterkornHelsaedGroenkorn

Bilag 2. Databeskrivelse















Formål

Formålet med dette notat er at beskrive hvordan data er klargjort til at kunne prediktere næste års afgrødevalg for en given mark.

Datainput

Data er leveret af Rita Hørfarter (RIH).

Nedfor ses hvilke filer der er brugt i denne analyse.

 AfgGrupper.xlsx	23-01-2023 14:21	Microsoft Excel-regne...	16 KB
 Marker_2011.xlsx	31-01-2023 14:31	Microsoft Excel-regne...	30.493 KB
 Marker_2012.xlsx	31-01-2023 15:45	Microsoft Excel-regne...	27.840 KB
 Marker_2013.xlsx	06-02-2023 15:50	Microsoft Excel-regne...	32.064 KB
 Marker_2014.xlsx	31-01-2023 14:25	Microsoft Excel-regne...	31.316 KB
 Marker_2015.xlsx	31-01-2023 14:27	Microsoft Excel-regne...	30.943 KB
 Marker_2016.xlsx	31-01-2023 14:28	Microsoft Excel-regne...	30.980 KB
 Marker_2017.xlsx	31-01-2023 14:28	Microsoft Excel-regne...	28.546 KB
 Marker_2018.xlsx	31-01-2023 14:29	Microsoft Excel-regne...	28.488 KB
 Marker_2019.xlsx	31-01-2023 14:30	Microsoft Excel-regne...	30.108 KB
 Marker_2020.xlsx	31-01-2023 14:30	Microsoft Excel-regne...	29.615 KB
 Marker_2021.xlsx	31-01-2023 14:25	Microsoft Excel-regne...	29.596 KB
 Marker_20220607_Beriget.xlsx	09-02-2023 13:10	Microsoft Excel-regne...	199.522 KB
 Master_MarkID2022_2011.xlsx	09-02-2023 13:05	Microsoft Excel-regne...	114.097 KB

I masterfilen "Master_MarkID2022_2011.xlsx" findes et unikt ID for hver mark. Her findes afgrøde historik samt et årsspecifikt mark-ID der bruges for at koble oplysninger om ejer cvr i de foregående år fra hver mark års-fil "Marker_xxxx.xlsx".

Master fil																	
ID	Areaal2022	Afgrøde2022	Afgrødekode2021	Afgrødekode2020	Afgrødekode2019	Afgrødekode2018	Afgrødekode2017	Afgrødekode2016	Afgrødekode2015	Afgrødekode2014	Afgrødekode2013	Afgrødekode2012	Afgrødekode2011	Afgrøde2022	Afgrødekode2022	Afgrødekode2021	
1	2,1	276	276	276	276	0	0	0	0	0	0	0	254	254	0	Permanent græs	Per
2	2,74	160	11	160	11	160	1	1	10	11	10	11	1	14	Sukkerroer til fabr	Vin	
5	17,1	1	160	11	22	1	11	11	22	11	11	1	108	Vårbyg	Sul		
6	0,21	276	251	251	251	251	252	252	310	310	310	310	310	Permanent græs	Per		
7	0,41	1	5	5	5	5	271	271	271	271	271	271	271	5	Vårbyg	Ma	

Master fil																
Afgrøde2022	Afgrødenavn2021	Afgrødenavn2020	Afgrødenavn2019	Afgrødenavn2018	Afgrødenavn2017	Afgrødenavn2016	Afgrødenavn2015	Afgrødenavn2014	Afgrødenavn2013	Afgrødenavn2012	Afgrødenavn2011	MarkID_2021	MarkID_2020	MarkID_2019	MarkID_2018	MarkID_2017
0	Permanent græs	Permanent græs	Permanent græs						Miljøgræs MVJ-til	Miljøgræs MVJ-til		529109	532895	0	0	0
4	Sukkerroer til fabr	Vinterhøvede	Vårbyg	Sukkerroer til fabr	Vårbyg	Vinterhøvede	Vinterhøvede	Vinterhøvede	Vårbyg	Røjgræs/ste. alm	Vårbyg	529663	532149	347746	279918	535712
8	Vårbyg	Sukkerroer til fabr	Vinterhøvede	Vinterhøvede	Vårbyg	Vinterhøvede	Vinterhøvede	Vinterhøvede	Vinterhøvede	Vinterhøvede	Vårbyg	529556	532399	346682	280276	533265
0	Permanent græs	Permanent græs	Permanent græs	Permanent græs	Permanent græs	Permanent græs	Permanent græs	Brak	Udyrkede landbru	Udyrkede landbru	Udyrket mark	575310	146189	124181	293316	140389
5	Vårbyg	Majs til modenhed	Majs til modenhed	Majs til modenhed	Majs til modenhed	Rekreative formål	Rekreative formål	Rekreative formål	Rekreative formål	Rekreative formål	Majs til modenhed	574370	146777	124725	292456	141010
0	Vårbyg	Permanent græs	Permanent græs	Græs under 50%	Græs under 50%	Græs med kløver	Græs med kløver	Græs med kløver	Udyrkede landbru	Udyrkede landbru	Udyrket mark	574644	146537	124976	292648	139725

MarkID					
MarkID_2016	MarkID_2015	MarkID_2014	MarkID_2013	MarkID_2012	MarkID_2011
0	0	0	188875	609657	0
12	175681	88110	452910	604699	6629
65	287503	494074	361703	428780	246527
89	471084	83654	487897	371691	467860
10	426144	462341	487674	371714	467821
25	76628	373468	487604	371732	467802
21	292038	36850	487646	371743	467838

Mark års-filerne findes for marker i årene 2011-2021.

ID	Journaln	Cvr	Marknr	lmk areal	Afgkode	Afgroede	Gb	Gbanmeldt	Aarstal
1	21-0027854	26022762	40-1	4,1	254	Miljøgræs MVJ-til	1	4,1	2021
2	21-0027854	26022762	40-2	1,2	254	Miljøgræs MVJ-til	1	1,2	2021
3	21-0027854	26022762	40-3	0,58	254	Miljøgræs MVJ-til	1	0,58	2021
4	21-0027854	26022762	40-4	2,91	907	Naturarealer, økol	0	0	2021
5	21-0027854	26022762	40-5	1,49	252	Permanent græs,	1	1,49	2021
6	21-0027854	26022762	50-0	0,32	257	Permanent græs,	1	0,32	2021

Info omkring blandt andet lokation og jbnr hentes fra "Marker_20220607_Beriget.xlsx" datafilen hvor info kobles på alle årene. Disse 2022 informationer antages her ikke at ændre sig, og kobles derfor på marken for alle årene.

ID	CVR	ID	CVR	Journaltitel	Marknr	Indtjningsha	Afgroede	Afgroede	Afsægt	Ghaenkel	Omdrift	IB	IB	Postby	Kommune	Kommune	Region	Kystvand	Kystvand	Overlap	Overlap	Retention	Retention	Retention	Afgroede	Afgroede		
1	76488255	22-00874	76488255	76488255	22-0087458	56-1	2,1	276	Permanenter	1 6831208	2,1	0	4	IB_5_9	4700	Næstved	370	Næstved	Region Sj	6,2E+07	25000013	Karrebæk	4,55E+05	0,002	47,4384	3,76308	44,7885	0
2	76488255	22-00874	76488255	76488255	22-0087458	85-0	2,74	190	Sukkerroer	1 689165-4	2,74	1	6	IB_5_9	4684	Holmestrap	370	Næstved	Region Sj	6,2E+07	25000013	Karrebæk	0,10546	0	72,7479	41,9329	53,9956	10
4	76488255	22-00874	76488255	76488255	22-0087458	34-0	17,1	1	Vårbyg	1 6911218	17,1	1	7	IB_5_9	4733	Tappermaj	370	Næstved	Region Sj	9,4E+07	29000020	Præste Fx	0	0	47,0241	14,4719	38,6327	11
5	76488255	22-00874	76488255	76488255	22-0087458	30-3	0,21	276	Permanenter	0 692119-5	0	0	6	IB_5_9	4733	Tappermaj	370	Næstved	Region Sj	9,4E+07	29000020	Præste Fx	0	0	97,0522	59,8657	83,8452	252
6	76488255	22-00874	76488255	76488255	22-0087458	39-10	0,41	1	Vårbyg	1 692119-5	0,41	1	6	IB_5_9	4733	Tappermaj	370	Næstved	Region Sj	9,4E+07	29000020	Præste Fx	0	0	28,941	0	27,1888	271
7	76488255	22-00874	76488255	76488255	22-0087458	18-9	1,01	1	Vårbyg	1 692119-5	1,01	1	6	IB_5_9	4733	Tappermaj	370	Næstved	Region Sj	9,4E+07	29000020	Præste Fx	0	0	47,0241	14,4719	38,6327	260

Der kobles afgrødegrupper på også. Disse grupperinger er manuelt lavet af Jacob Krog og der er genereret en excel-fil "AfgGrupper.xlsx" der indlæses og bruges til denne gruppering.

Afgkode	AfgNavn	GrpKode	GrpNavn
1	1 Vårbyg	Grp_1	Vaarsaed
3	3 Vårhavre	Grp_1	Vaarsaed
4	2 Vårhvede	Grp_1	Vaarsaed
5	7 Korn + bælgssæd under 50% bælgssæd	Grp_1	Vaarsaed
6	4 Blanding af vårsæde arter	Grp_1	Vaarsaed
7	6 Vårhvede, brødhvede	Grp_1	Vaarsaed
8	55 Vårrug	Grp_1	Vaarsaed
9	56 Vårtriticale	Grp_1	Vaarsaed
10	8 Vårspelt	Grp_1	Vaarsaed
11	58 Sorghum	Grp_1	Vaarsaed
12	216 Silomajs	Grp_216	Majs
13	5 Majs til modenhed	Grp_216	Majs
14	11 Vinterhvede	Grp_11	Vinterhvede
15	13 Vinterhvede, brødhvede	Grp_13	Brødhvede
16	15 Vinterhybridrug	Grp_15	Vinterrug
17	14 Vinterrug	Grp_15	Vinterrug
18	16 Vintertriticale	Grp_15	Vinterrug
19	10 Vinterbyg	Grp_10	Vinterbyg
20	9 Vinterspelt	Grp_9	VinterkornOvr
21	17 Blanding af efterårssæde arter	Grp_9	VinterkornOvr

Dataproces

Her i udviklingsfasen er dataprocesen beskrevet som nedenfor. Når der skal laves det endelige modul til virkemiddelvalgeren rettes dette til.

Der hentes først data via "*DataInput_versx.R*". Derefter laves der datamanagement i "*Datamanagement_versx.R*". Datamodeller testes af i separate scripts efterfølgende.

Dataredigering

Cvr:

Vi vil kun se på bedrifter der er ejet med cvrnummer. Der sorteres i hvilke cvrnumre i data der er gyldige og som ikke indikerer de er personejet. Cvrnr overskrives med 0 hvis:

- Indeholder xxxx
- Missing
- Ikke består af 8 cifre
- Eller er lig 00000000

Marker/observationer er fjernet hvis:

- Ingen cvrejer i perioden 2011-2022
- Hvis 2022 ikke er cvrejet
- Hvis cvrnummer i 2021 er ukendt, da vil 2022 være første afgrøde obs for bedriften.

OBS med disse sorteringer fjerner jeg marker/ID hele vejen igennem ikke bedrifter.

Det vil sige. Jeg kan have bedrifter der ikke er fuldstændige med alle deres marker, da nogle vil være fjernet grundet ovenstående kriterier.

Sekvenser:

Casen er at kunne forudsige næste års afgrødevalg. Bruge denne prædiktion som sand og med den inkluderet igen forudsige året efter.

Med vores datagrundlag er casen her at "være" i år 2020 og ville prædiktere afgrødevalg i 2021. Samt bruge denne prædiktion til at prædiktere 2022 afgrødevalget. For at vi ved hvad der er valgt i 2021 og 2022 til modeludvikling og vi har så valide data som muligt kræver vi at:

- Cvrnr 2022 er kendt.
- Hvis forrige cvrnr i sekvensen er ukendt, da fjernes observation og vi medtager kun de kendte observationer fra bedriften.
- Kræver 4 eller flere observationer i sekvensen.
Dette vil medføre at når 2022 og 2021 fjernes og bliver dem der skal prædikteres, da vil vi stadig have minimum 2 observationer fra bedriften eller 1 obs fra bedriften samt 1 obs fra tidligere ejers afgrødevalg (2020 samt 2019 data).

Afgrødegrupper:

Gruppering af afgrødekode er lavet af Jakob Krog. Se Appendix for inddeling pr. 11/4-2023.

Gruppenavnet brugt her for en samling af afgrøder er koden på den mest udbredte afgrøde i gruppen.

Det er kun medtaget marksekvenser hvor:

- Afgrødegruppen er kendt i 2022.

Der er besluttet at der skal fjernes de bedrifter som har marker med mere end 1 ha af gartneri, plante eller frugtbar (koder grp_420, grp_489, grp_510).

Da pyntegrønt Juletræer og energiafgrøde Lavskov (koderne grp_582 og grp_519) har en omdriftstid på 10 år er der valgt at fjerne de marker hvis koderne er i 2020, 2021 eller 2022 data, da vil det ikke give mening at prædiktere for 2021, da det vil være samme kode, og derfor forstyrrer de.

Hvis der historisk har været koderne i år 2011 op til 2020 da overskriver vi koderne med NA, da vi gerne vil have marken og dens historik med efter der har været juletræer.

Marker der i hele perioden har registreret AreallkkeRetur (kode grp_318) fjernes helt.

Hvis "Ukendt navn" (kode grp_44) findes i 2013 eller frem da er marken fjernet. Dette er forholdsvis få observationer. Der findes dog en del marker med observationer af koden i 2011 og 2012. I disse tilfælde overskrives koden med missing, og markerne med den efterfølgende afgrødesekvens indgår da i modellen.

Disse ovenstående filtreringer gør at vi får marker med afgrøde sekvenser uden huller.

Bemærk at afgrøden "Kartofler" ikke er håndteret hensigtsmæssigt i databehandlingen. Da de første krav til datainput lagde vægt på bedriftens historik, da blev kartoffelmarkerne med andet cvrnummer set som et "ejerskifte", og indgår derfor ikke nødvendigvis rigtigt. Dette er der ikke gået mere i dybden med her. Men der er et ønske om dette håndteres anderledes i efterfølgende analyser, her er tanken at overskrive cvrnummeret når kartofler er der, med samme cvrnummer, så det indgår i bedriftens sekvens alligevel.

Bedriftstyper:

Bedriftstyperne er bestemt ud fra afgrødegrupperne.

Hver mark bliver defineret som

- "Vinter" hvis afgrødegruppenavnet indeholder "Vinter".
- "Kvaeg" hvis afgrødegruppenavnet indeholder "Fodergraes" eller "Majs".
- "Froe" hvis afgrødegruppenavnet indeholder "Frograes", "Spinat" eller "Havefrø".
- "Other" hvis afgrødegruppenavnet indeholder alt andet end det ovenstående.

Derefter ses der på hvor stor en andel af arealet de forskellige ovenstående grupper udgør af det samlede areal på bedriften. I overskrivende rækkefølge er det nedenstående bedriftsareal der afgør bedriftstypen. Så vi siger bedriftstypen er

- "Froe" hvis arealandelen af "Froe" > 15.
- "Kvaeg" hvis arealandelen af "Kvaeg" > 50.
- "Vinter" hvis arealandelen af "Vinter" > 50.
- "Other" alt der ikke kan defineres ved ovenstående.

Trænings-/validerings- og testdata

Casen er at "stå" i 2020 og ville prædiktere afgrødevalg i 2021. Samt bruge denne prædiktions til at prædiktere 2022 afgrødevalget.

Der indgår altså alle datasekvenserne fra 2011-2020 i træningsdata. Da 2021 er kendt, da bruges hele 2011-2021 som test til at validere op i mod på prædiktionserne. Samme for 2022.

Dataopgørelser

I masterfilen findes 574.04 unikke mark observationer for 30.354 unikke cvrnumre (Obs dette inden cvr-redigering, så den er ikke helt præcis). Af disse marker er 17.114 med missing cvrnummer i 2022.

Efter cvr redigering samt mark redigering, da er vi på 556.926 unikke marker og 30.353 unikke 2022-cvrnumre.

Efter sekvens selektion samt de sidste redigeringer da har vi et slut træningsdatasæt på ses der på 366.051 unikke marker, på 24.793 unikke 2022-cvrnumre.