



Notat v. 2017: Konkursmodeller for de danske landbrugserhverv

22-12-2017

Philipp Trénel, ph.d., seniorkonsulent, specialist
phtr@teknologisk.dk, +45 72 20 33 92

Dette notat er fortroligt og underligger vilkårene beskrevet i bilaget "Fortrolighedsaftale.pdf".

Formål

- At udvikle version 2.0 af en driftsgren-specifik (søer, slagtesvin, planteavl, kvægdrifter) model til prædiktion af konkurssandsynlighed på baggrund af data fra SEGES økonomi-database (ØDB) og konkursdata indhentet af SEGES. Modellen er videreudviklet ud fra modellen beskrevet i notat af samme navn fra 2016 (se bilag 1). Konkursmodellen sigter på at indgå i SEGES early-warning system rettet mod rettidig rådgivning af landmænd med henblik på forebyggelse af konkurser.

Overordnede konklusioner

- Den her anvendte algoritme på baggrund af Random Forest modeller kan med rimelig høj nøjagtighed (accuracy = [0.84, 0.87]) prædiktere en bedrifts konkurssandsynlighed på baggrund af DLBRs ØDB data.
- Denne prædiktionsnøjagtighed er relativt stabil også i forecast-situationer op til 5 år frem.
- Der vurderes derfor, at der er grundlag for en videreudvikling af et early-warning system.
- Den lave prævalens af konkurser har betydning for, hvor mange konkurs-prædikterede bedrifter, der rent faktisk kommer til at gå konkurs (PPV-værdi). PPV'en kan øges på bekostning af sensitiviteten til en identifikation af højrisiko-bedrifter (f.eks. PPV ~ 0.75 og sensitivitet ~ 0.3 for kvægbedrifter).
- Der er mange udfordringer forbundet med en Big Data analyse, som den her gennemførte. En succesfuld/brugbar model skal imødekomme en lang række udfordringer, såsom prævalens, data kvalitet og missing data, features extraction, variable selection, honest test set based validation, tolkning.
- Se også bilag [her](#).

Metoder

- Metoder følger notatet fra 2016, men med følgende ændringer/tilføjelser:
 - A. Tiltag til forbedring af tolkning af konkursmodellen med fokus på den enkelte bedrift
 1. Gruppering af ØDB input-variabler
 2. Funktioner til beregning af bedrift-specifik variabel importance (= variabler der bidrager til en øget konkurssandsynlighed)
 3. Forslag til præsentation af en bedrifts model-prædikterede konkurssandsynlighed og tilhørende prædiktions-usikkerhed.
 - B. Tiltag til forbedring af prædiktionsnøjagtigheden i konkursmodellen
 1. Nye data fra 2016 og 2017
 2. Nye ØDB input-variabler, deriblandt bank (Fig. 1), konjunkturindikatorer, m.m.
 3. Model på tværs af driftsgrene, men med driftsgrenspecifik forecast.
 4. Kontekstuel analyse: alle variabler udtrykkes både som populationsniveau-variabler (driftsgren- og årsspecifikt gennemsnit) og som bedriftsniveau-variabler relativt til populationsniveau-variablen (udtrykt som percentil), se Fig. 2.
 5. Fler-trins modeller: en sammensat model, der arbejder i to steps: Step 1 prædikterer konkurssandsynligheden; step 2 genprædikterer konkurssandsynligheden for de bedrifter fra step 1 med en prædikteret konkurssandsynlighed $> p$.
 6. Inkludering af tidligere års prædikterede konkurssandsynlighed som input-variabel i model til forecast af fremtidig konkurssandsynlighed.
 7. Imputering: Da modellen forventes at køre i et on-the-fly IT-system er der implementeret en ny imputeringsstrategi (missing value imputering) baseret på en uni-variater tidsserie-tilgang (Stineman-interpolering).
 8. Parameter tuning: forbedring af random forest modellens prædiktions grundet fine-tuning af model-parametre såsom antal træer, antal variabler pr. træ, m.m.

Resultater

- **A1** Gruppering af ØDB input-variabler:
Der foreligger redundant information i ØDB data grundet variabernes indbyrdes korrelationer, hvilket har en potential negativ effekt på modellernes prædiktionsperformance. Derudover begrænses fortolkningen af modellens prædiktions af de mange variabler. Derfor er der udført en analyse af input-variabernes indbyrdes korrelationer og en hierarkisk cluster analyse af disse (cluster tærskel $r = 0.75$). På baggrund af dette har SEGES grupperet variablerne i omtrent 10 klasser for hver driftsgren, se Fig. 2 og 3 og bilagene.

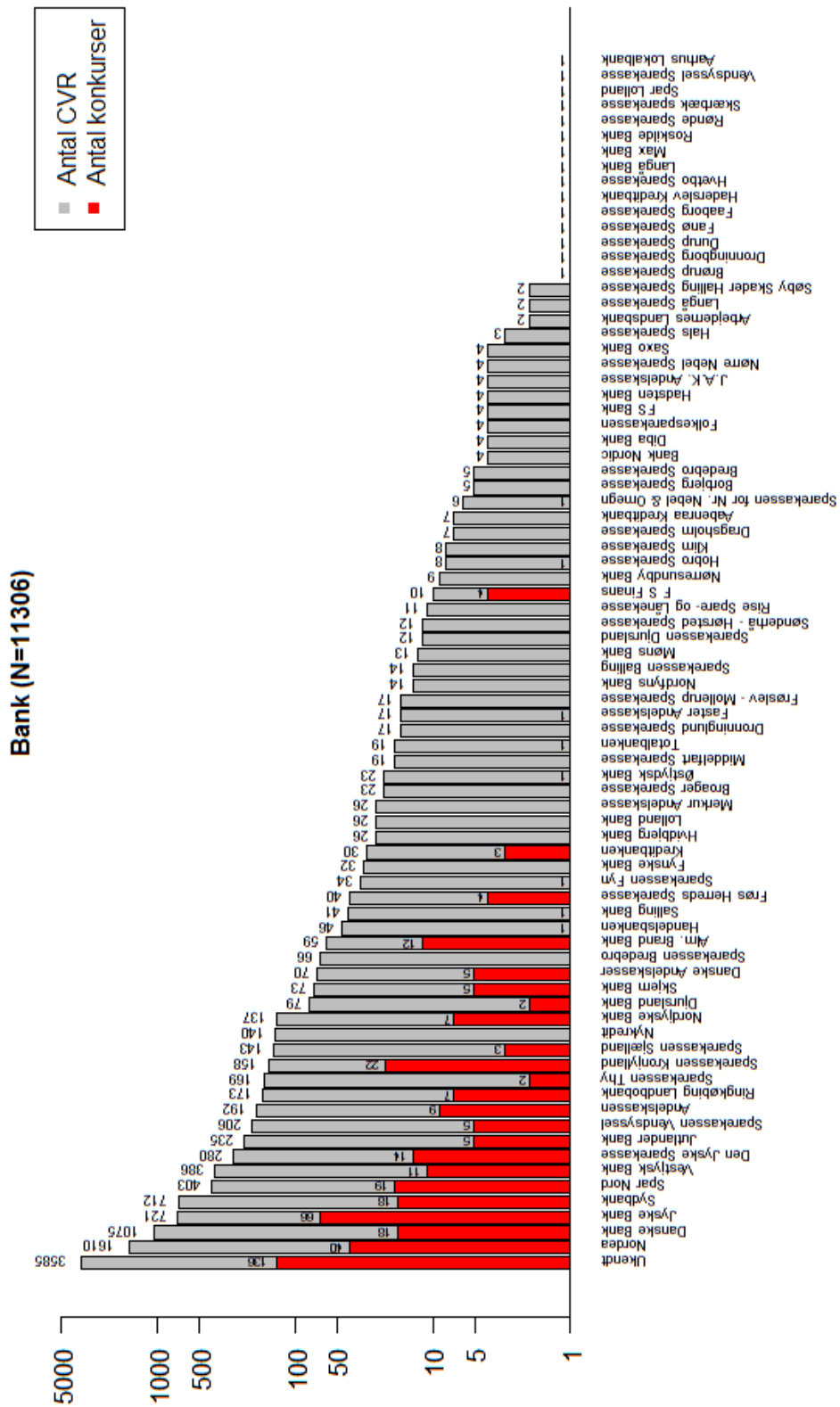


Fig. 1: Bank-variablen.

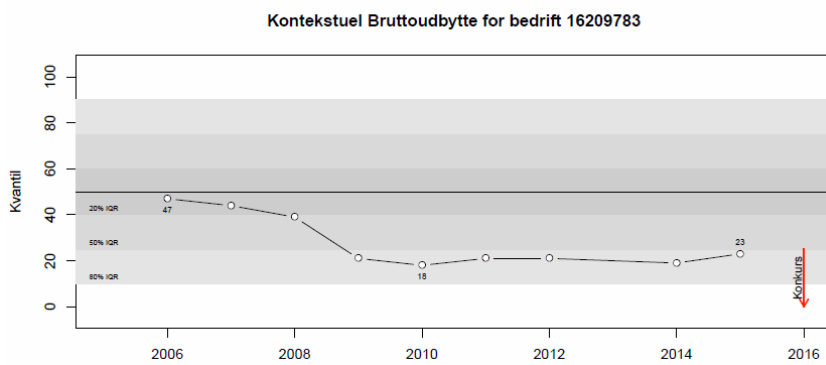
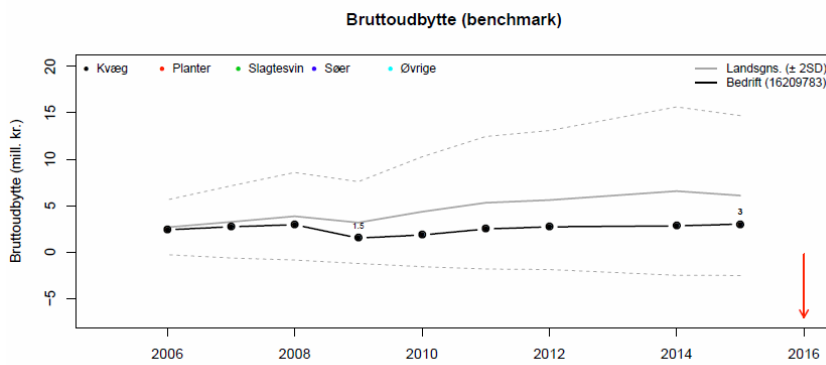
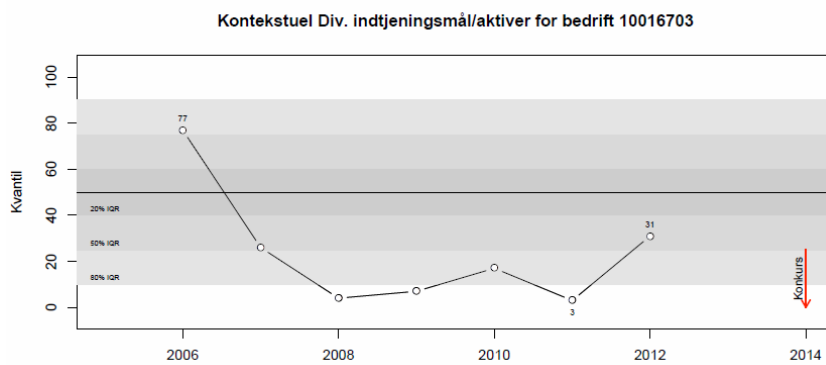
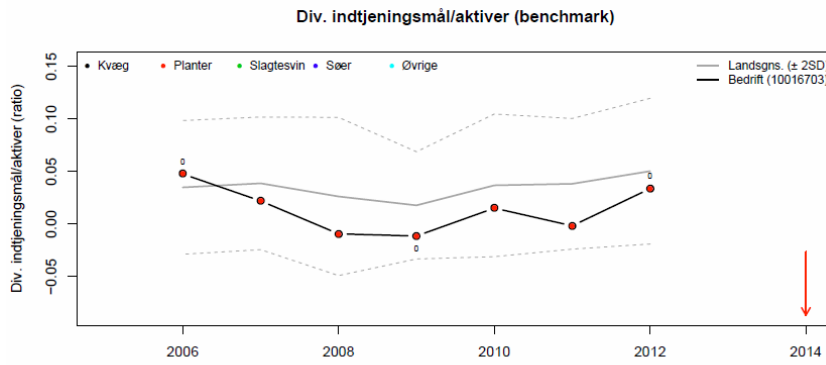


Fig. 2: Den kontekstuelle analysetilgang eksemplificeret ved to bedrifter.

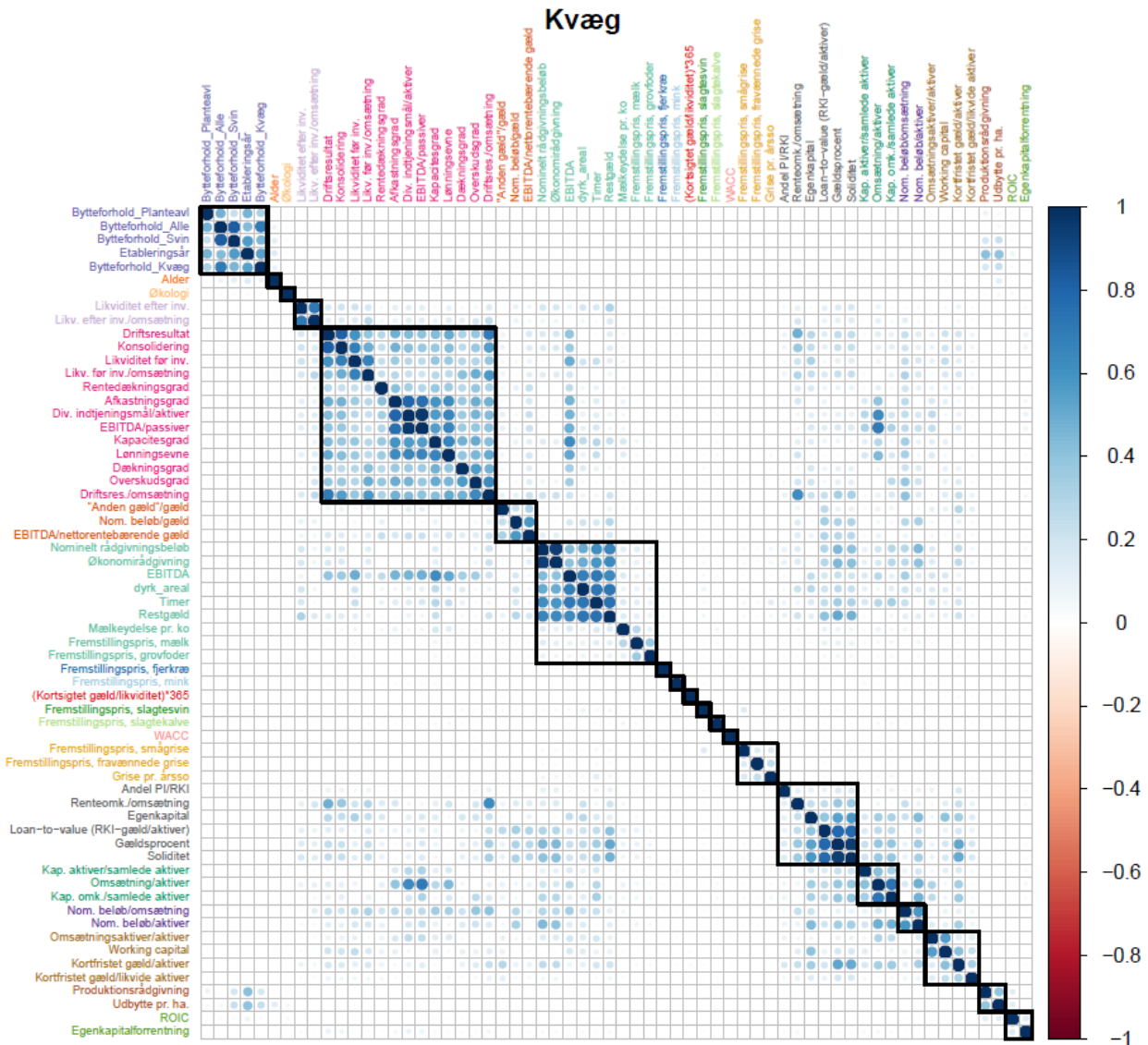


Fig. 2: ØDB inputvariablenes indbyrdes Pearson korrelation for kvæg. De fra den hierarkiske cluster analyse resulterende clusters af højt korrelerede variabler ($r > 0.75$) er angivet som sorte bokse.

Gruppering af mulige forklarende variable i ranket rækkefølge – Mælkeproducenter (75 pct.)

Rentabilitet, likviditet, indtjening mv. Driftsresultat/omsætning Omsætning/aktiver Likv. før inv./omsætning Driftsresultat (nominelt beløb) Renteomk./omsætning Rentedækningsgrad Kapacitetsgrad Dækningsgrad Omsætning (nominelt beløb) EBITDA/passiver Likviditet efter inv. Likviditet før inv. Lønningsevne Div. Indtjeningsmål/aktiver Afkastningsgrad Likv. efter inv./omsætning EBITDA (nominelt beløb) Overskudsgrad Konsolidering (ny)	Finansielt: Egenkapitalforrentning ROIC	Soliditet mv. Gældsprocent Soliditet Kortfristet gæld/aktiver Egenkapital Andel PI/RKI Kortfristet gæld/likvide aktiver EBITDA/nettorentebærende gæld Aktiver (nominelt beløb) Restgæld (nominelt beløb) Working capital (kortf. Aktiver/kortf.gæld) Loan-to-value (RKI-gæld/aktiver) Kap.omk./samlede aktiver Gæld (nominelt beløb) *Anden gæld*/gæld Kap.aktiver/samlede aktiver Omsætningsaktiver/aktiver	Produktivitet: Udbytte pr. ha. Produktionsrådgivning	Rådgivning og ejendomsstørrelse: Rådgivning (nominelt beløb) Økonomirådgivning Dyrket areal Timer (ny)
	Finansielt beredskab: (Kortsigtet gæld/likviditet)*365		Effektivitet: Mælkeydelse pr. ko Fremstillingspris pr. kg. mælk Fremstillingspris, grovfoder Fremstillingspris, Slagtekalve (ny)	
	Finansieringsomkostninger: WACC		Markedsforhold: Bytteforhold, alle (Ny) Bytteforhold, mælk (Ny)	

Fig. 3: Gruppering af ØDB inputvariabler for kvæg.

- **A2** Funktioner til beregning af bedrift-specifik variabel importance:
 Der er udviklet R-procedurer til at trække følgende informationer ud af modellens skov for hver bedrift:
 - En prædikeret konkurssandsynlighed, $Pr(Y=1)$
 - En PPV
 - Variabelgruppernes (kategoriernes) betydningsgrad (BG i %), dvs. bidrag til $Pr(Y=1)$, i rangeret orden
 - Variabelgruppernes gns. forskel til $Pr(Y=1) = 0.5$ tærsklen
 - Variablernes betydningsgrad BG (i %) indenfor deres variabelgruppe, dvs. bidrag til gruppens BG, i rangeret orden
 - Variablernes gns. forskel til $Pr(Y=1) = 0.5$ tærsklen
 - Grafer over variablernes effekt på $Pr(Y=1)$ (kun hovedvirkning)
- **A3** Forslag til præsentation af en bedrifts model-prædikerede konkurssandsynlighed og tilhørende prædiktions-usikkerhed:

Se boks på næste side og Fig. 4 som baggrund.

	højeste risikogruppe		forhøjet-risiko-gruppen		risiko-gruppen		lav-risiko gruppen	
Pr(Y = 1)	(1 - 0.718]		(0.718 - 0.556]		(0.556 - 0.398]		(0.398 - 0]	
	Fra	Til	Fra	Til	Fra	Til	Fra	Til
Acc.	0.84	0.87		0.91		0.93		0.91
Se	0	0.20		0.50		0.80		0.91
Sp	1	0.99		0.99		0.987		0.91
PPV ^a	-	0.91		0.77		0.53		0.40

a: Ved 6% 5-års periode-prævalens

A) Vedr. **Kunde A**, med en model-prædikeret **konkurssandsynlighed på 0.8**.

- Vi vil tilbyde dig rådgivning, idet SEGES konkursmodel indikerer, at din besætnings konkurssandsynlighed er meget høj og estimeret til at være højere end 89% af besætningerne, der forventes ramt af konkurs indenfor det næste år. Konkursprædiktionen for besætninger med en estimeret konkurssandsynlig lig eller større end din har en sikkerhed på over 99%¹. Du er dermed i den **højeste risikogruppe**.
- Beregning: $1 - Se = 0,89$ ved en tærskel på 0.8; PPV 95%-konfidensinterval = (0.99, 1) ved en tærskel på 0.8 og prævalens=0.06

B) Vedr. **Kunde B**, med en prædikeret **konkurssandsynlighed på 0.6**.

- Vi vil tilbyde dig rådgivning, idet SEGES konkursmodel indikerer, at din besætnings konkurssandsynlighed er forhøjet og estimeret til at være højere end 61,9% af besætningerne, der forventes ramt af konkurs indenfor det næste år. Konkursprædiktionen for besætninger med en estimeret konkurssandsynlig lig eller større end din har en sikkerhed på 83%¹. Du er dermed i **forhøjet-risiko-gruppen**. Vi vil gerne hjælpe dig med at udpege fokusområder med størst potentiale til omgående at sænke din konkurssandsynlighed.
- Beregning: $1 - Se = 0,619$ ved en tærskel på 0.6; PPV = 0,833 ved en tærskel på 0.6 og prævalens=0.06

C) Vedr. **Kunde C**, med en prædikeret **konkurssandsynlighed på 0.4**.

- SEGES konkursmodel indikerer, at din besætnings konkurssandsynlighed ikke er signifikant lavere end 50%. Du er dermed i **risiko-gruppen**. Vi kan tilbyde dig et rettidigt økonomisk eftersyn for at identificere nøgleparametre, der bidrager betydeligt til din konkurssandsynlighed.
- Beregning: Se = 0,80 og Sp = 0.90 ved en tærskel på 0.4; PPV = 0,4997, c.i. = (0,467; 0,531) ved en tærskel på 0.4 og prævalens=0.06

D) Vedr. **Kunde D**, med en prædikeret **konkurssandsynlighed på 0.1**.

- SEGES konkursmodel indikerer, at din besætnings konkurssandsynlighed er lav. 91% af besætninger med en konkurssandsynlighed på niveau med din eller lavere forventes ikke at gå konkurs. Dermed er du i **lav-risiko gruppen**. Såfremt du alligevel ønsker økonomisk rådgivning, er du velkommen til at kontakte os.
- Beregning: Se = 0,91 og Sp = 0.91 ved en tærskel på 0.1; PPV = 0,396 ved en tærskel på 0.1 og prævalens=0.06
- Du gøres opmærksom på, at SEGES konkursmodellens prædikterede konkurssandsynlighed er en indikator med ingen garant for hverken konkurs eller ikke-konkurs. Formålet med SEGES konkursmodellen er at hjælpe moderat til stærkt konkurstruede besætninger vha. økonomisk rådgivning. Modellens opgave er at udpege risiko-bedrifter tids nok til en rettidig omhu kan effektueres.

¹: ved en antaget 5-års periode-prævalens på konkurser på 6%.

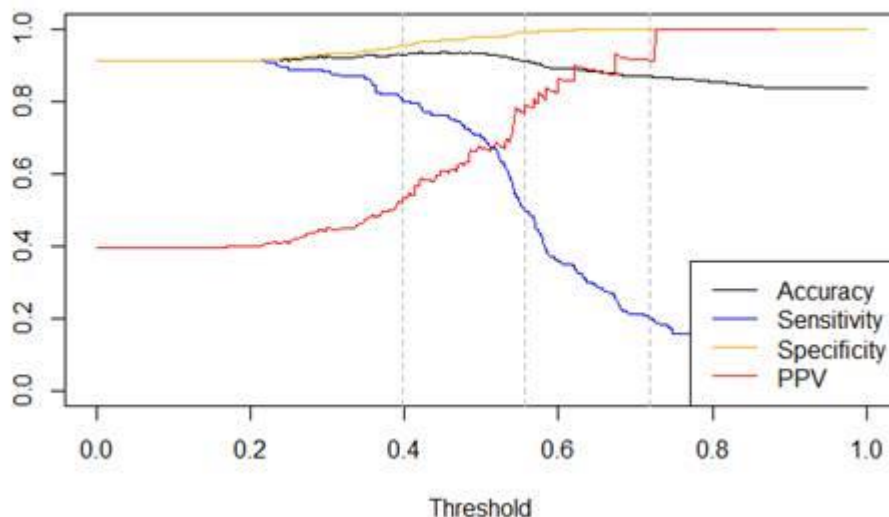


Fig. 4: Evaluering af sikkerheden i konkurs-prædiktionen vist som sammenhæng mellem Accuracy, Sensitivity, Specificity og Positive Predictive Value (PPV) i forhold den valgte konkurssandsynligheds-threshold for én af de kørte modeller. Grafen viser, at sikkerheden i prædiktionen af konkurs (PPV) stiger med konkurssandsynligheden. Tilsvarende stiger sikkerheden af prædiktionen af ikke-konkurs med faldende konkurssandsynlighed (NPV, ikke vist).

Tabel 1: Fundne prædiktionsperformance-forbedringer ved forskellige tiltag udført i 2017. Prædiktionsperformance for trin 1 til 3 er vist for driftsgren kvæg.

Trin/tiltag	inkluderede trin ¹	Prædiktionsperformance	
		Sensitivitet	Specificitet
0. Version 1 (2015 og før)	0	0.88	0.87
1. Nye data (2016 + 2017)	0,1	0.83	0.83
2. Nye variabler	0,1,2	0.86	0.86
3. Kontekstuel analyse	0,1,2,3	0.87	0.87
4. Trans-driftsgren model	0,1,2,3,4	0.90	0.90
5. 2-steps model	0,1,2,3,4,5	0.90	0.90
6. Rullende model	0,1,2,3,4,6	0.91	0.91

¹: Stineman-imputering og parameter tuning er inkluderet i alle modeller

- **B Tiltag vedr. forbedring af prædiktionsperformance:**
 - Inkludering af nye data fra 2016 (og for konkursbedrifters vedkommende også 2017) har – under anvendelse af modellen fra 2016 – resulteret i et generelt fald i prædiktionsnøjagtigheden udtrykt som sensitivitet og specificitet, se tabel 1.
 - De nye tiltag afprøvet i 2017 har kunnet løfte prædiktionsperformance, resulterende i en forventet gennemsnitlig prædiktionsfejl på 9%.
 - Det konkluderes derfor, at inkludering af variabler, der beskriver den generelle konjunktur, både i form af bytteforhold (som nye variabler) og i form af den kontekstuelle analysetilgang, forbedrer prædiktionsperformance.
 - To-step modellen har ikke øget prædiktionsnøjagtigheden betydende.
 - Den rullende modelstrategi har marginalt løftet prædiktionsnøjagtigheden, men kræver yderlig implementering, såsom forbedret imputering.