



NTNU

Kunnskap for en bedre verden

Tilstandsmodellering av avløpsrør ved bruk av maskinlæring

Evaluering av modelltyper og deres mulighet til å generaliseres for norske kommuner.

Joakim Skjelde (NTNU) – joaskj@stud.ntnu.no / joakimskjelde99@gmail.com

Franz Tscheikner-Gratl (NTNU)

Marius Møller Rokstad (NTNU)

Bardia Roghani (NTNU)

Shamsuddin Daulat (NTNU)

Bakgrunn og mål

Bakgrunn

- Flere kommuner mangler data for å etablere maskinlæringsmodeller
- Enormt investeringsbehov i vann- og avløpssektoren

Mål (Forskningsspørsmål)

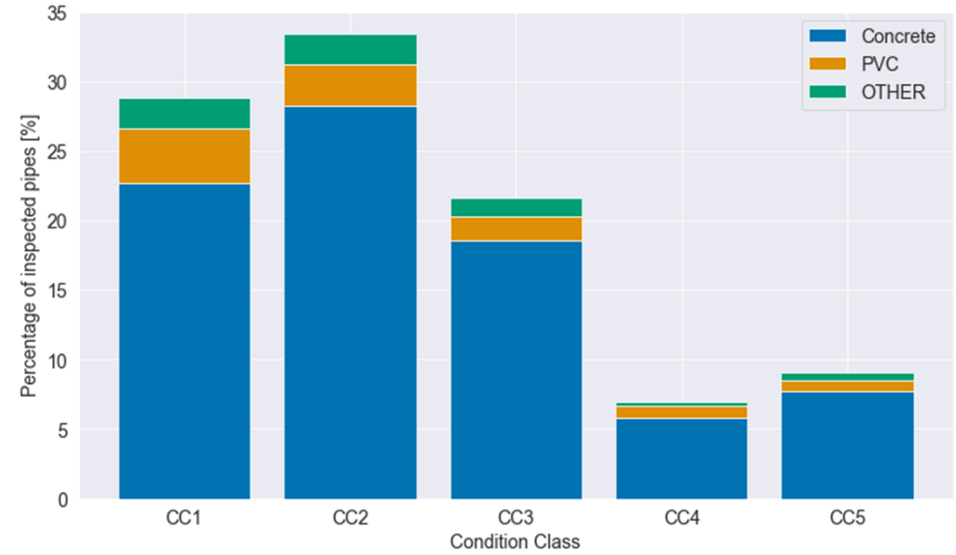
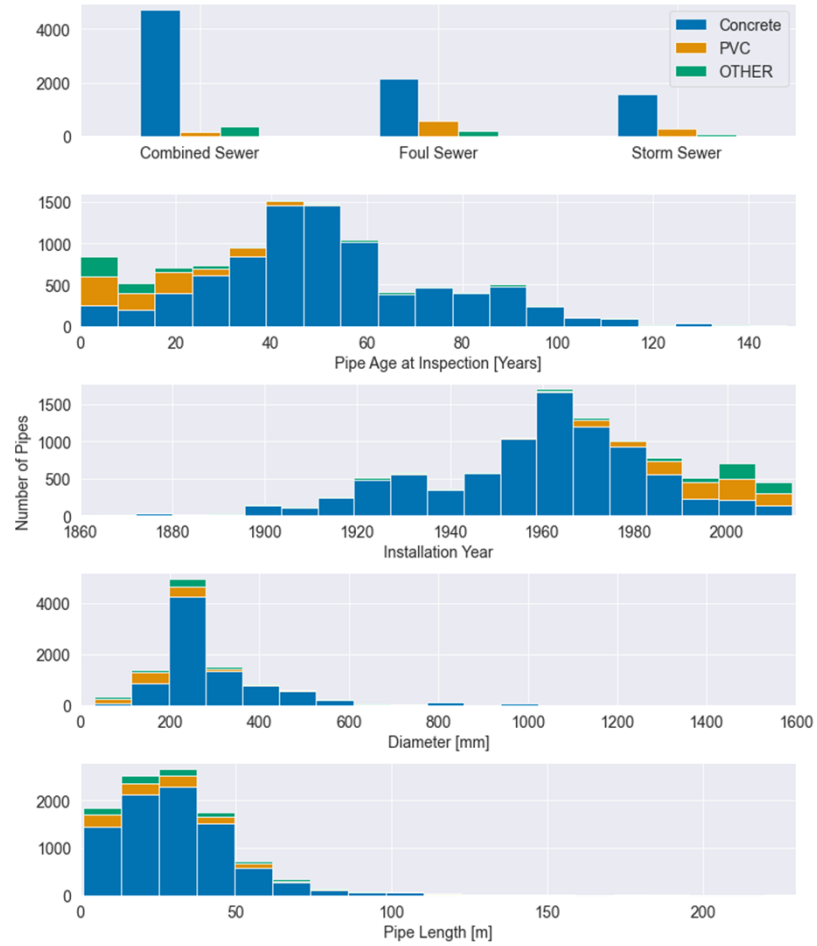
- Kan en «Global» model trent på flere kommuner brukes til å forutse rørtilstand i en annen kommune?
- Hvordan presterer en lite brukt modelltype i forhold til mer etablerte modeller?
- Hvilke parametere påvirker nedbrytningen av avløpsrør i denne studien?

Data

- Rørinspeksjoner fra 5 Norske kommuner
- Totalt ~10000 inspeksjoner, med lengde ~315km
- Fem forklaringsvariabler brukt
 - Alder
 - Lengde
 - Diameter
 - Type (AF, OV, SP)
 - Material (Betong, PVC, Annet)
- Mye manglende data!



Data



Municipality	Number of pipes	CC1	CC2	CC3	CC4	CC5
		(%)	(%)	(%)	(%)	(%)
1	6435	27	35	23	6	9
2	263	72	26	2	0	0
3	586	23	33	22	8	14
4	1251	31	38	20	5	6
5	1645	28	26	21	13	12
Total	10180	29	33	22	7	9

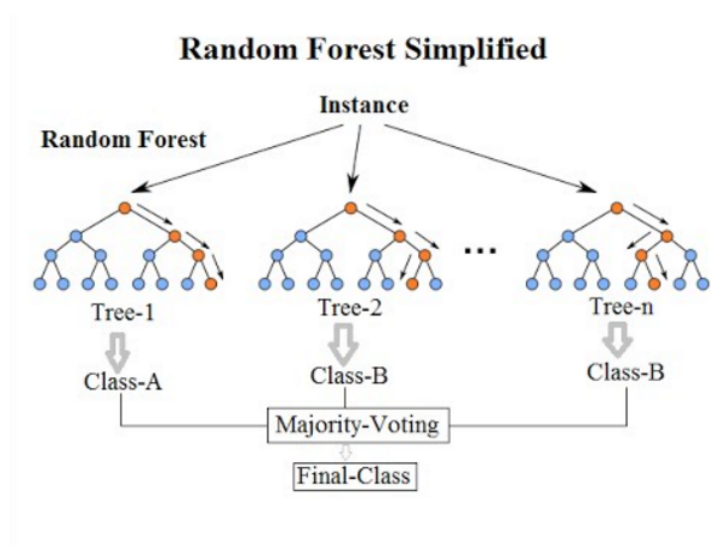


Metoder

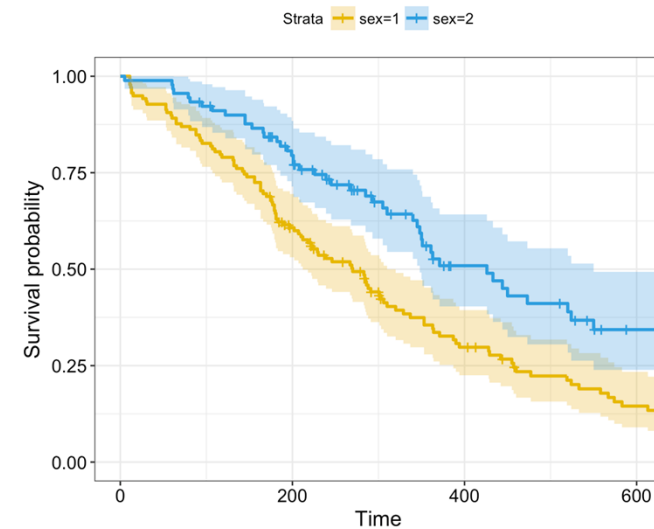
- Overførbarhet (Transferability)
 - Leave-One-Out (LOO)
 - Testes for to maskinlæringsmodeller
 - Overlevelsesmodell
 - Klassifiseringsmodell
- Sammenligning (Comparability)
 - Sammenligning av overlevelsesmodellen på to nivåer
 - Nettverksnivå, mot en statistisk modell
 - Rørnivå, mot en klassifiseringsmodell

Random Survival Forest (RSF)

Random Forest



Survival Analysis



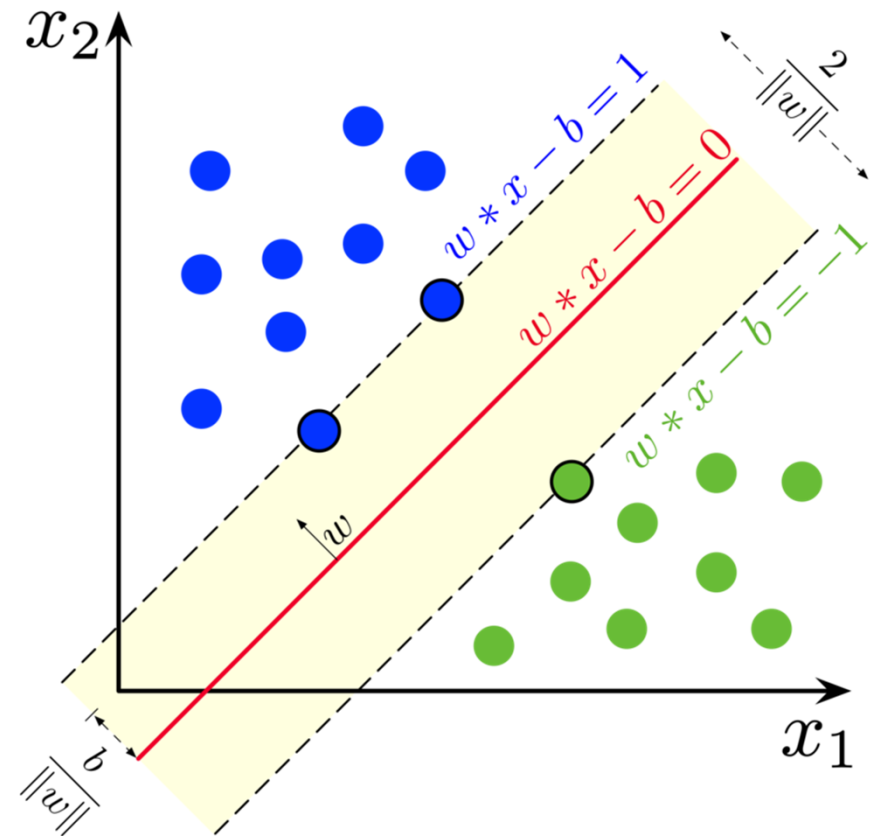
En kombinasjon mellom Random Forest og Survival Analysis

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *Annals of Applied Statistics*, 2(3)

Support Vector Machine (SVM)

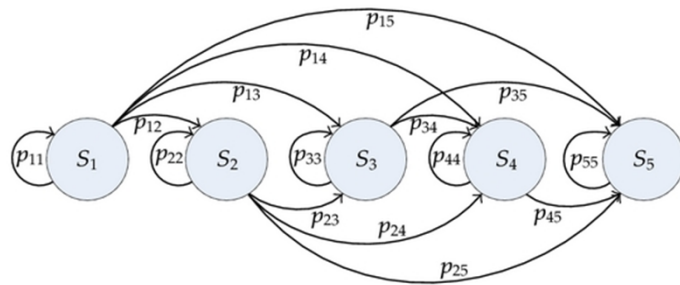
- Algoritme for klassifisering
- Basert på geometrisk avstand og bruk av «kernel»-funksjoner

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3)

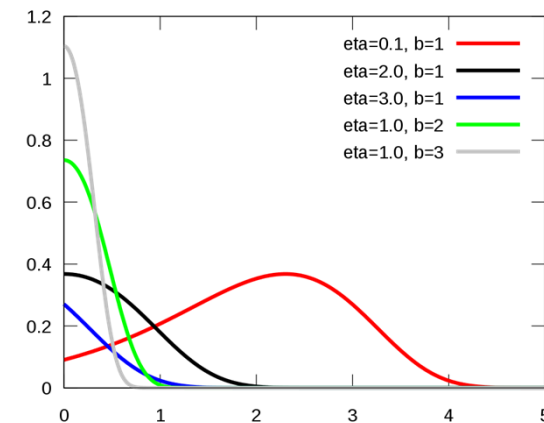


GompitZ

- Statistisk modell basert på Markov-kjeder og Gompertz sannsynlighetsfordeling
- Modellerer sannsynlighet for å bli i eller bytte tilstandsklasse



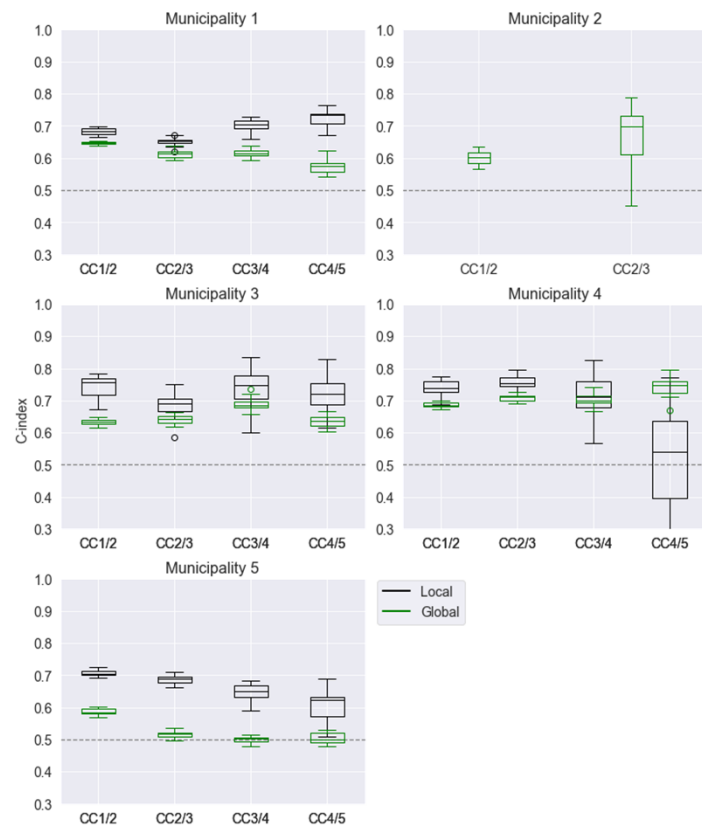
$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ 0 & p_{22} & p_{23} & p_{24} & p_{25} \\ 0 & 0 & p_{33} & p_{34} & p_{35} \\ 0 & 0 & 0 & p_{44} & p_{45} \\ 0 & 0 & 0 & 0 & p_{55} \end{bmatrix} \text{ where } p_{ij} \geq 0 \text{ and } \sum_j p_{ij} = 1$$



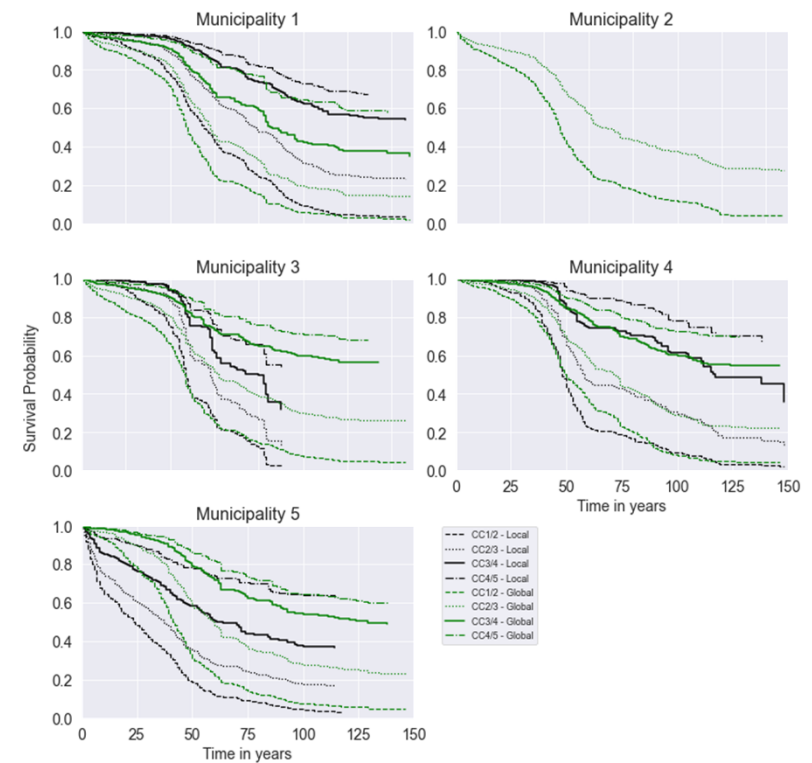
Le Gat, Y. (2008). Modelling the deterioration process of drainage pipelines. *Urban Water Journal*, 5(2)

Resultater: Overførbarhet (RSF)

Concordance Index

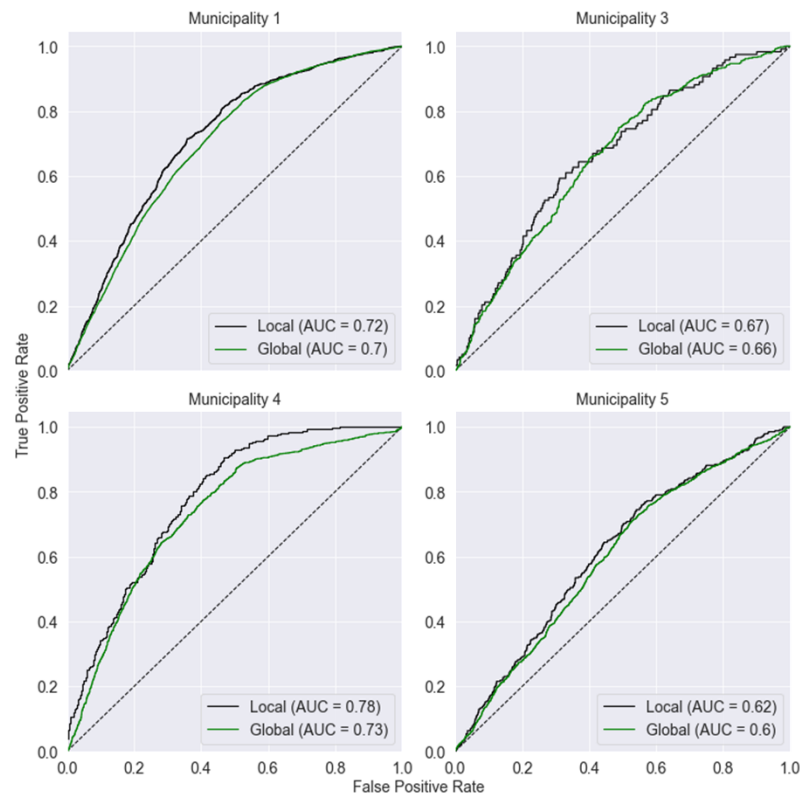


Overlevelseskurver



Resultater: Overførbarhet (SVM)

ROC-AUC



Usikkerhetsmatrise

		Predicted Condition (Local)				
		CC1	CC2	CC3	CC4	CC5
Actual Condition	CC1	37	19	5	6	6
	CC2	29	49	11	10	12
	CC3	3	15	10	8	4
	CC4	2	3	0	3	4
	CC5	0	1	2	3	9

		Predicted Condition (Global)				
		CC1	CC2	CC3	CC4	CC5
Actual Condition	CC1	27	15	13	8	10
	CC2	30	30	27	11	13
	CC3	10	7	6	8	9
	CC4	4	6	1	1	0
	CC5	4	4	2	3	2

		Predicted Condition (Local)	
		Good (CC1-3)	Bad (CC4-5)
Actual Condition	Good (CC1-3)	178	46
	Bad (CC4-5)	8	19

		Predicted Condition (Global)	
		Good (CC1-3)	Bad (CC4-5)
Actual Condition	Good (CC1-3)	165	59
	Bad (CC4-5)	21	6

Resultater: Overførbarhet (SVM)

Usikkerhetsmatrise

		Predicted Condition (Local)				
		CC1	CC2	CC3	CC4	CC5
Actual Condition	CC1	37	19	5	6	6
	CC2	29	49	11	10	12
	CC3	3	15	10	8	4
	CC4	2	3	0	3	4
	CC5	0	1	2	3	9

		Predicted Condition (Global)				
		CC1	CC2	CC3	CC4	CC5
Actual Condition	CC1	27	15	13	8	10
	CC2	30	30	27	11	13
	CC3	10	7	6	8	9
	CC4	4	6	1	1	0
	CC5	4	4	2	3	2

		Predicted Condition (Local)	
		Good (CC1-3)	Bad (CC4-5)
Actual Condition	Good (CC1-3)	178	46
	Bad (CC4-5)	8	19

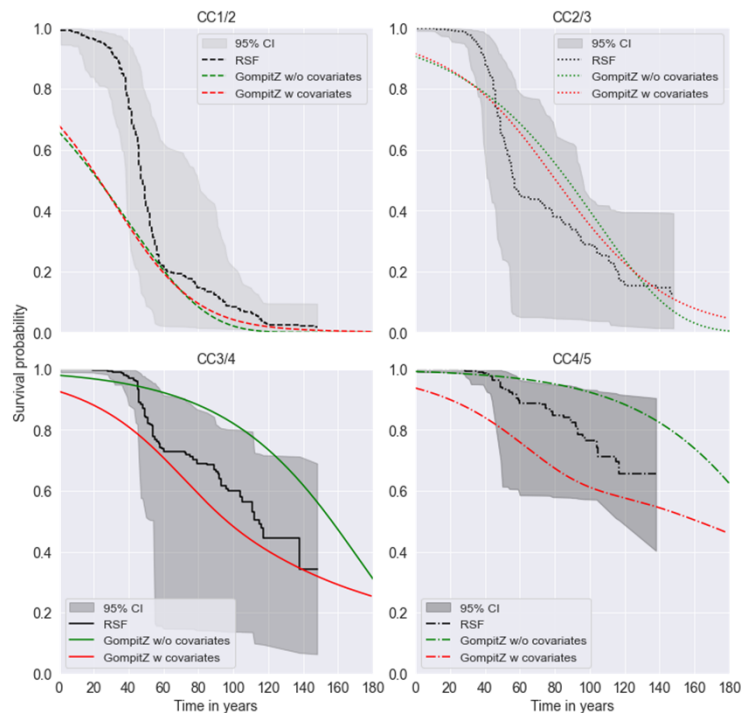
		Predicted Condition (Global)	
		Good (CC1-3)	Bad (CC4-5)
Actual Condition	Good (CC1-3)	165	59
	Bad (CC4-5)	21	6

Treffsikkerhet

- Fem klasser
 - Lokal: 43.0%
 - Global: 26.3%
- To klasser
 - Lokal: 78.5%
 - Global: 68.1%

Resultater: Sammenligning

Nettverksnivå

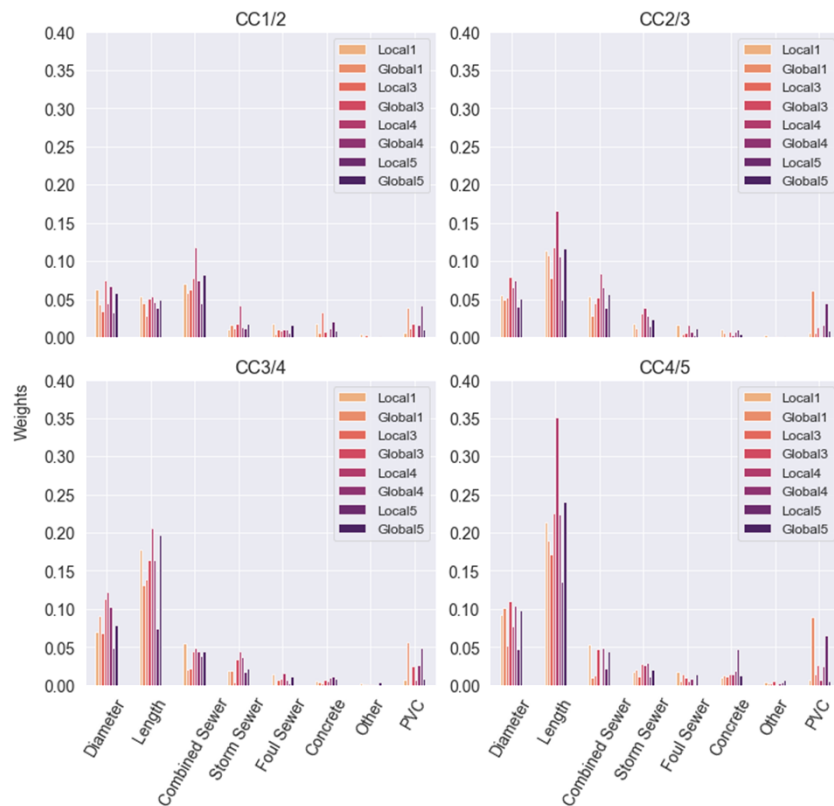


Rørnivå

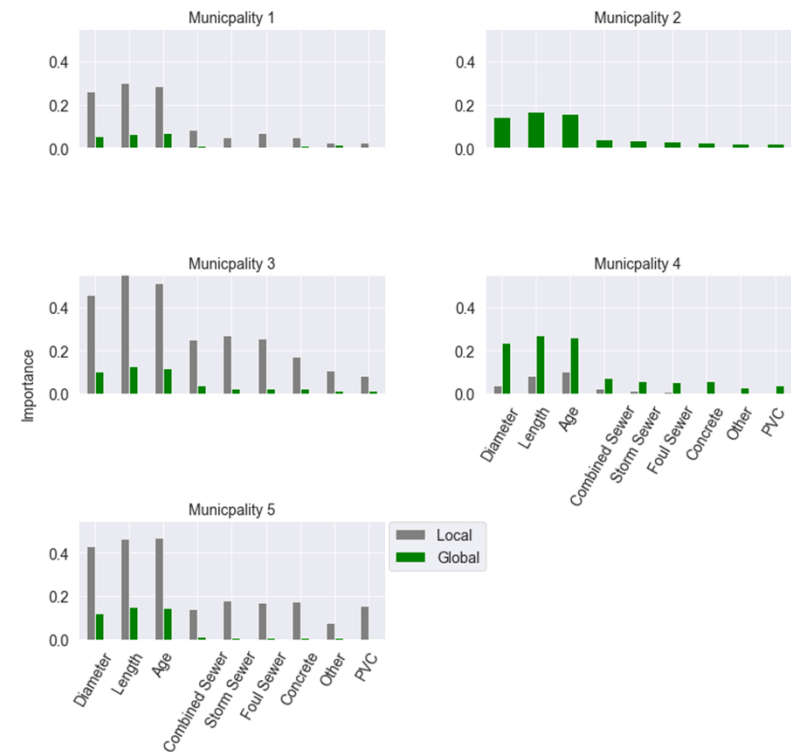
		SVM	
		Predicted Condition	
		Good	Bad
Actual Condition	Good	178	46
	Bad	8	19
Accuracy = 78.5%			
$K_{pipe} = 0.240$			
RSF with probability cutoff = 0.50		Predicted Condition	
		Good	Bad
Actual Condition	Good	221	3
	Bad	24	3
Accuracy = 89.2%			
$K_{pipe} = 0.513$			
RSF with probability cutoff = 0.88		Predicted Condition	
		Good	Bad
Actual Condition	Good	176	48
	Bad	6	21
Accuracy = 78.5%			
$K_{pipe} = 0.217$			

Resultater: Forklaringsvariable

Random Survival Forest



Support Vector Machine





Til ettertanke

- Data er ett «must» innen maskinlæring
 - Men dataen må også være rett
- Tilstandsmodeller kan bli forflyttet mellom representative kommuner
- Fem klasser gjør modellene mindre treffsikre, redusere til to eller tre?
- Det er fortsatt noe ukjent rundt nedbrytningsprosessen
 - Lengden er viktigst!
 - Treffsikkerhet har sjelden vært over 80%, selv med mange forklaringsvariabler

Takk for oppmerksomheten!

- Ta kontakt om du har spørsmål rundt oppgaven, funnene, eller muligheter og utfordringer med tilstandsmodellering generelt
- Joakim Skjelde (NTNU) – joakskj@stud.ntnu.no / joakimskjelde99@gmail.com