





# Decision Matrix

 Valg	 Beskrivelse	 Column	 Dato
<u>Hvilket metode skal vi velge for å analysere dataserien?</u>	Vi har tre kategorier: Statistisk modellering (Guassian mixture model, K-means clustering, random forest), temporal feature modeling (LSTM, RNN) og spatial feature modeling (CNN, DNN)		
<u>Snevre inn scope på antall kategorier</u>	Etter samtale med veileder så fant vi ut at 1300 kategorier kan være litt stort for to masterstudenter, og at det kan være lurt å snevre inn dette antallet for å få et mer overkommelig problem for en masteroppgave. Problemet er at mange kategorier antageligvis korrelerer en del med hverandre, og ved å se på hver enkelt individuelt så får vi ikke det hele bildet. I en perfekt verden så burde vi gjort en covarians analyse av alle kategoriene og laget "clusters" med de kategoriene som korrelerer med hverandre.		@September 27, 2021
<u>Spissing av oppgaven mot anomaly detection</u>	Vi stod mellom cluster analysis og, anomaly detection. I dialog med veileder og prisguiden gikk vi for anomaly detection fordi: A. Det gav mye utbytte for prisguiden. De får stort utbytte av å finne ut av en trend tidlig, istedenfor 1.5 mnd etter det har tatt av. B. Veileder virket gira		@August 31, 2021
<u>Vi bestemte at vi bruker ARIMA som baseline</u>	Sammen med Anders bestemte vi at vi bruker ARIMA som baseline. Vi må fortsatt gjøre research for å bygge et argument for dette		@October 11, 2021

 Valg	 Beskrivelse	 Column	 Dato
<u>Går bort ifra anomaly detection over til anomaly prediction</u>	Anders mente at bare ren anomaly detection er et lite spennende problem som alt har blitt løst. Det er mye mer spennende å prøve å spå et anomaly, og når det kommer prøve å si noe om hvordan det vil utvikle seg.		@October 11, 2021
<u>Velger metode CNN autoencoder LSTM</u>	Dette er noe som ikke er blitt prøvd på innen vårt domene. Det virker kult... TODO skriv emr		@October 11, 2021
<u>Vi la inn e-commerce og sales i search termer.</u>	Vi fant ut at vi hadde veldig få papers som gikk på time series prediction generelt, spesielt innenfor domene vårt. E-commerce og salg tidsserier kan tenkes at følger samme distrubisjon som vårt dataset, med trender, perioder og mye støy.		@October 25, 2021
<u>Changed problem P to be more generic</u>	Fjernet "in user click events" fra problemet, da det ekskluderer veldig mange problemstillinger	<u>Structured Literature Review</u>	@November 1, 2021