



NETCOMPANY

Workshop - Maskinlæring

VERSJON

1.0

FORFATTER

Henrik Bøhler

Ole Kristian Stumpf

Petter Fagerlund Asla

Øystein Kvamme Repp

Hva skal vi gå gjennom?

- Del 1
 - i. Introduksjon til maskinlæring
 - ii. Bruksområder
 - iii. Nevrale nettverk
- Del 2
 - i. Oppgaver

Artificial Intelligence (AI)

Simulert menneskelig intelligens

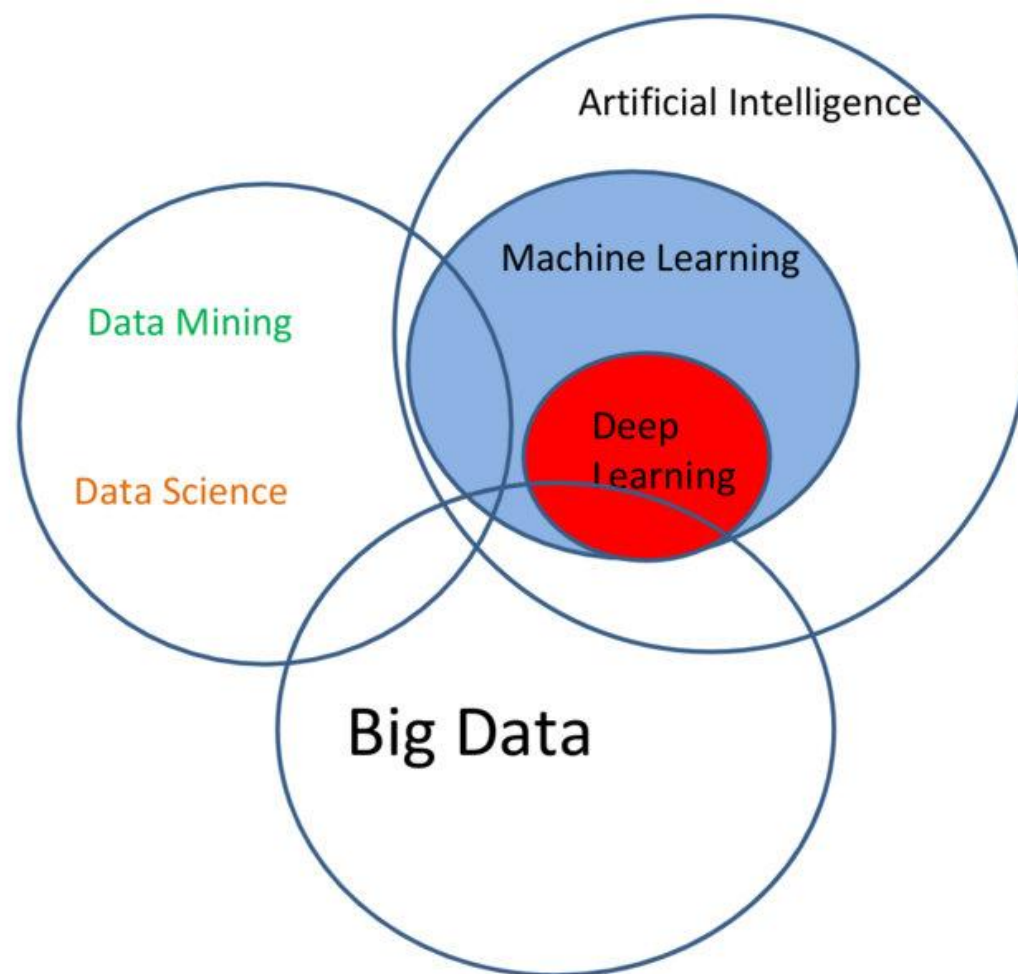
- Lære
- Resonere
- Løse problemer
- Prosessere informasjon

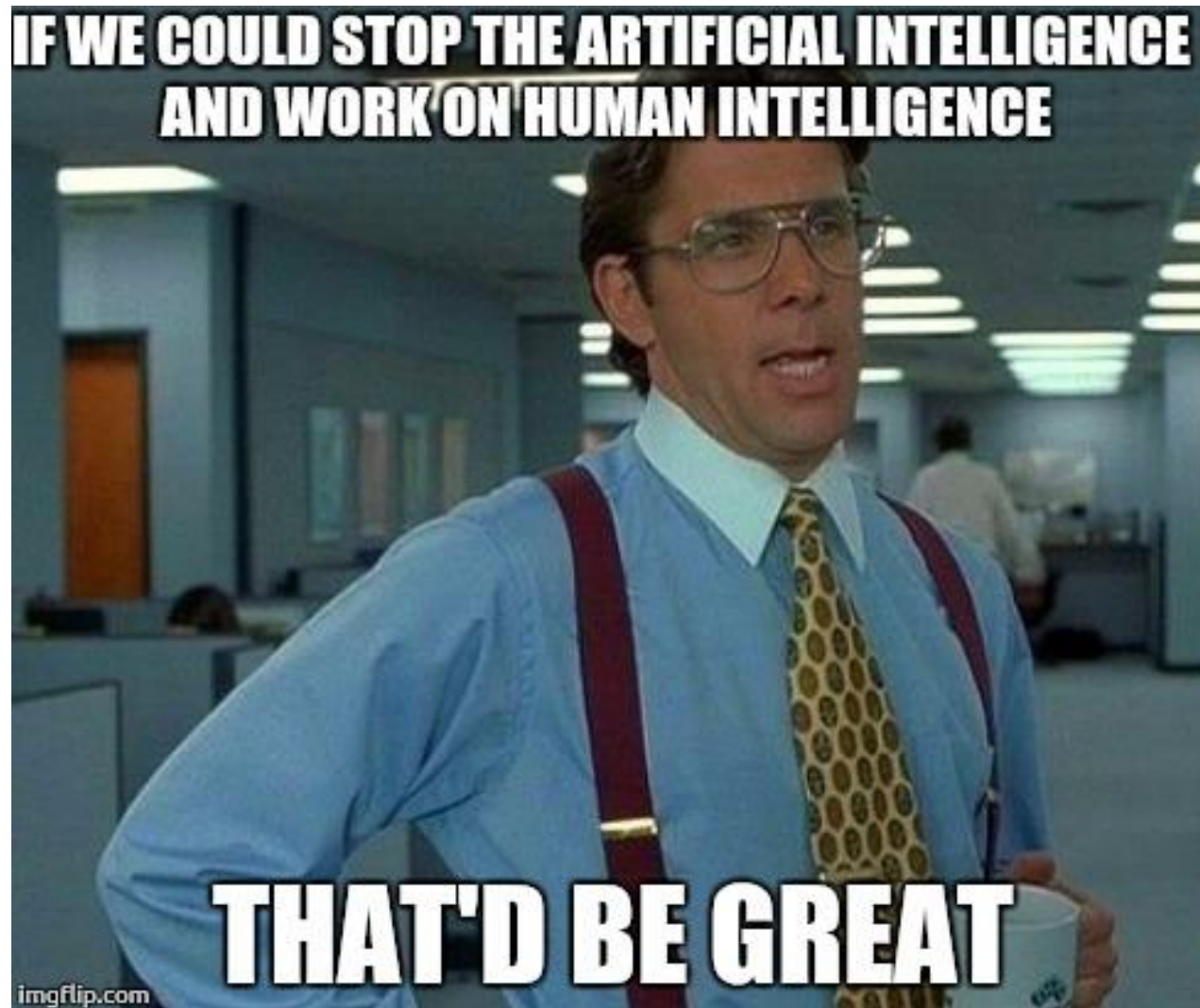
The
Economist

Artificial intelligence

Anything you can do, AI can do better.
So how will it change the workplace?

<http://learnmore.economist.com/story/57a849c338ba0ee26d98a68d>





Hva er maskinlæring?

Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed

- [Arthur Samuel](#) (1959)

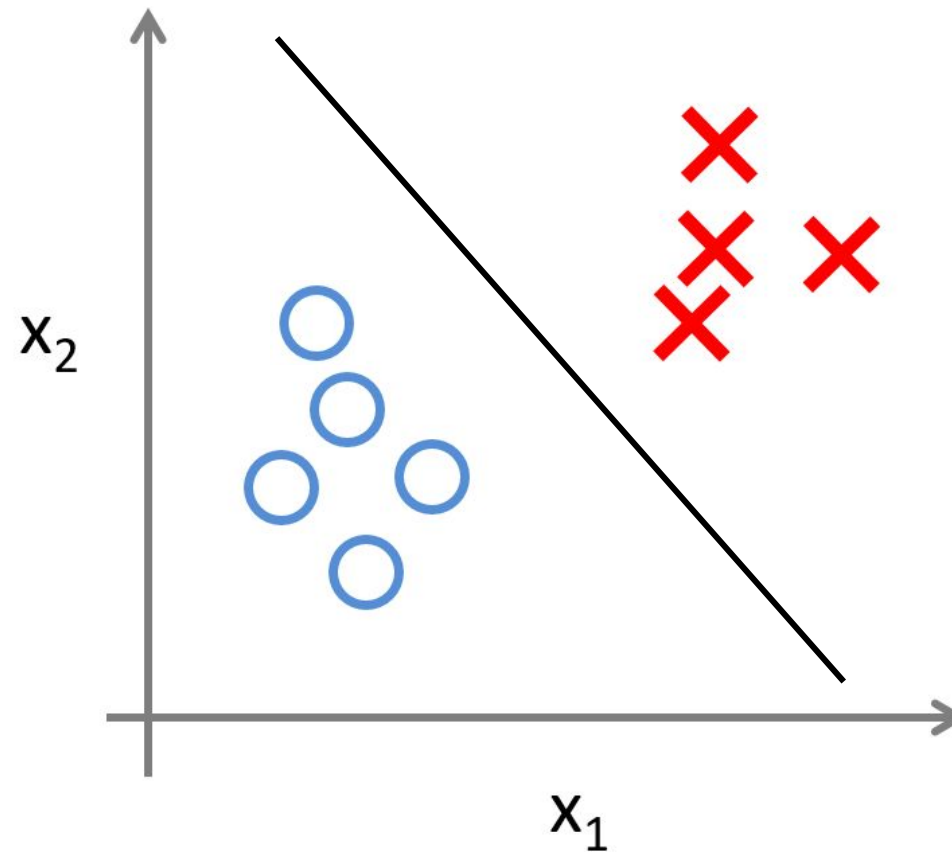
Hva er maskinlæring?

- Implementasjon av AI
- Algoritmer som lærer
- Ofte store mengder data - mer jo bedre!
 - Finne mønster
 - Predikere det ukjente basert på det kjente

Hvilke problemer kan maskinlæring løse?

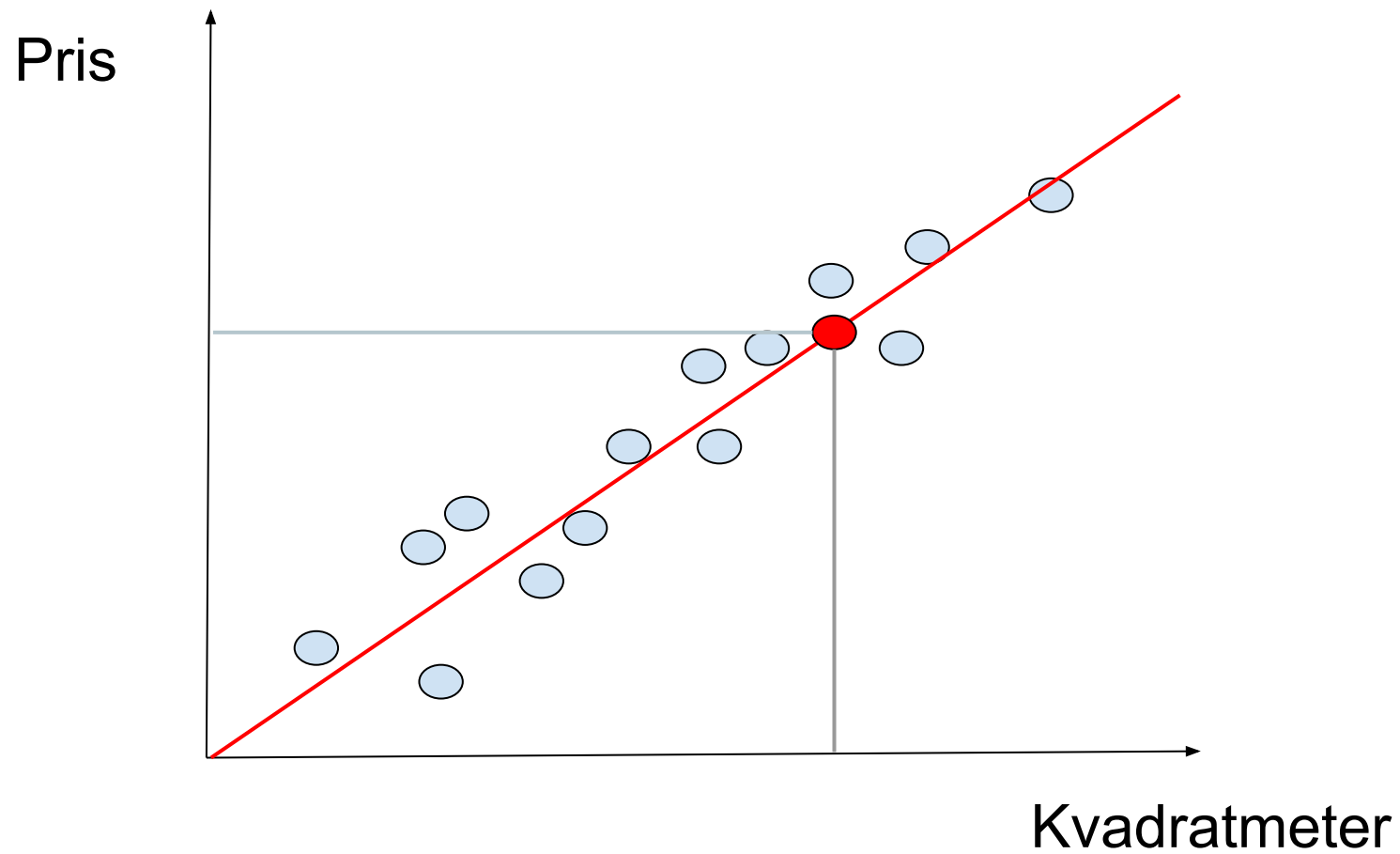
Automatisk søknadsvurdering

Supervised - klassifisering



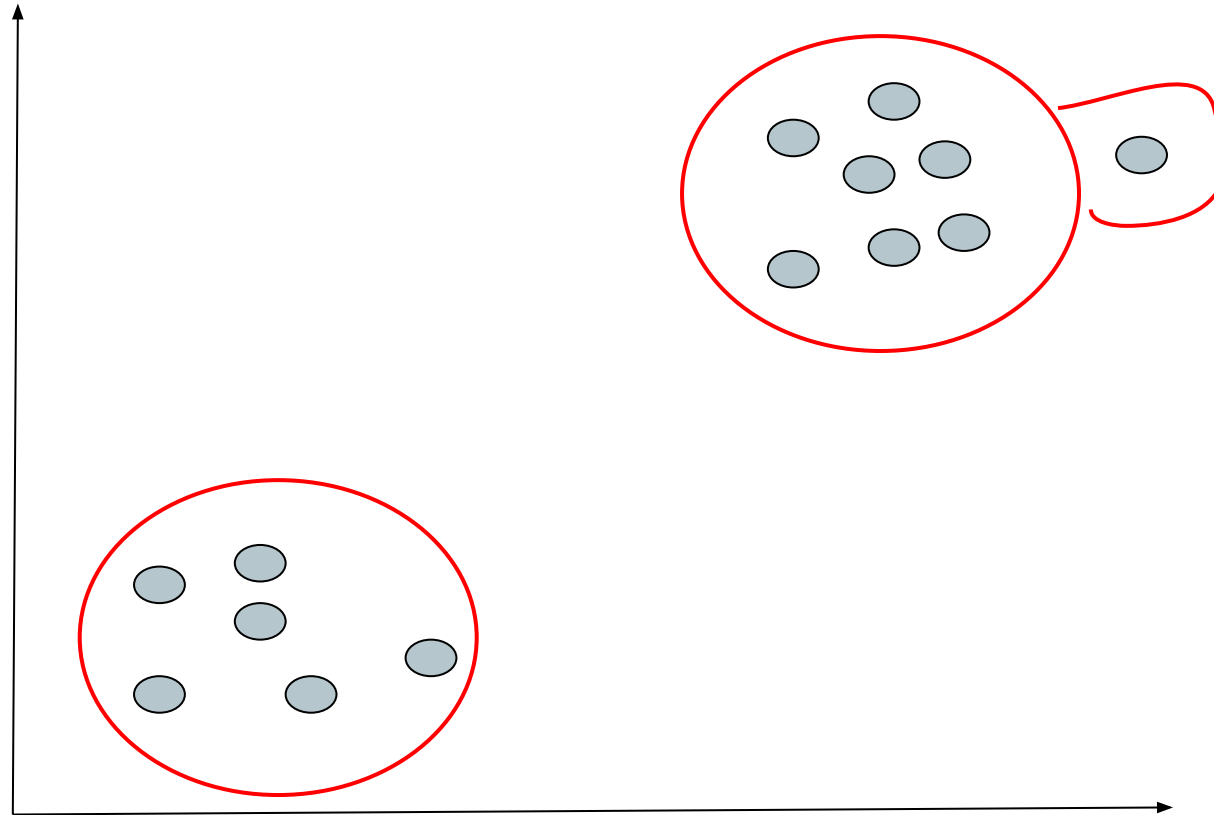
Hva er forholdet mellom
kvadratmeter og pris på bolig?

Supervised - regresjon



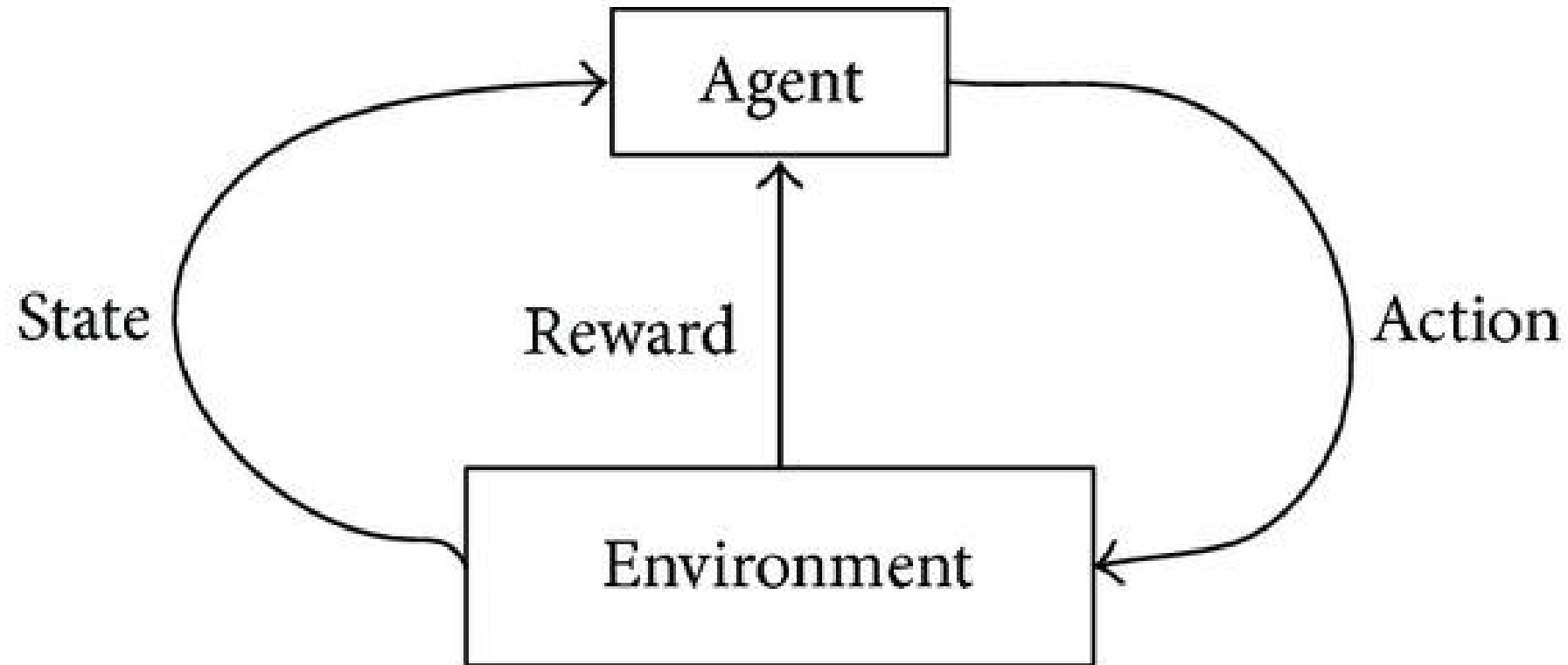
Musikkenbefaling

Unsupervised - clustering



Annonser

Reinforcement learning



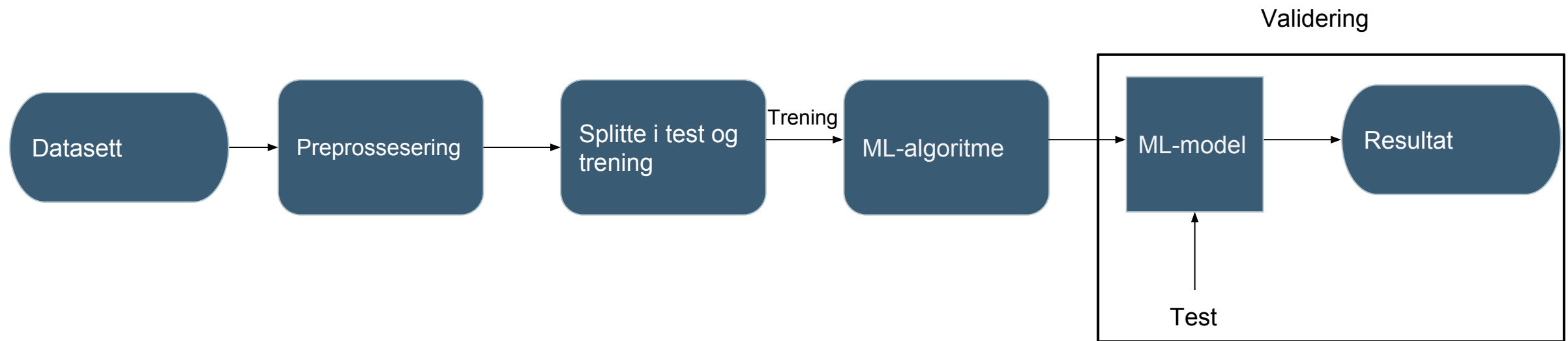


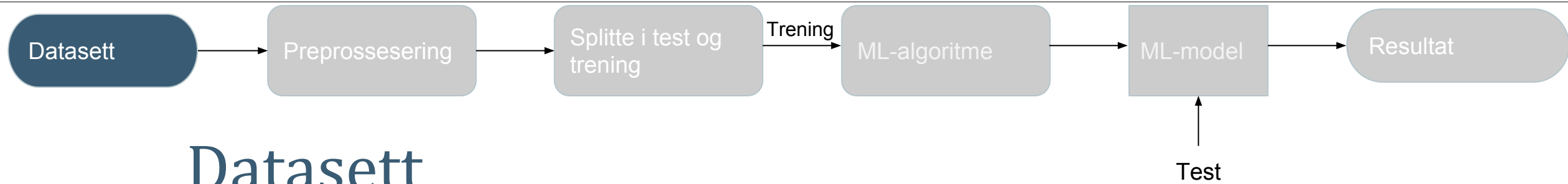
Netcompany og maskinlæring



<https://marketoonist.com/wp-content/uploads/2014/04/140414b.correlation.jpg>

Læringsprosessen





Datasett

- Kan være hva som helst
- Setter føringer for alle videre steg

Datasett

Preprossesering

Splitte i test og
trening

Trening

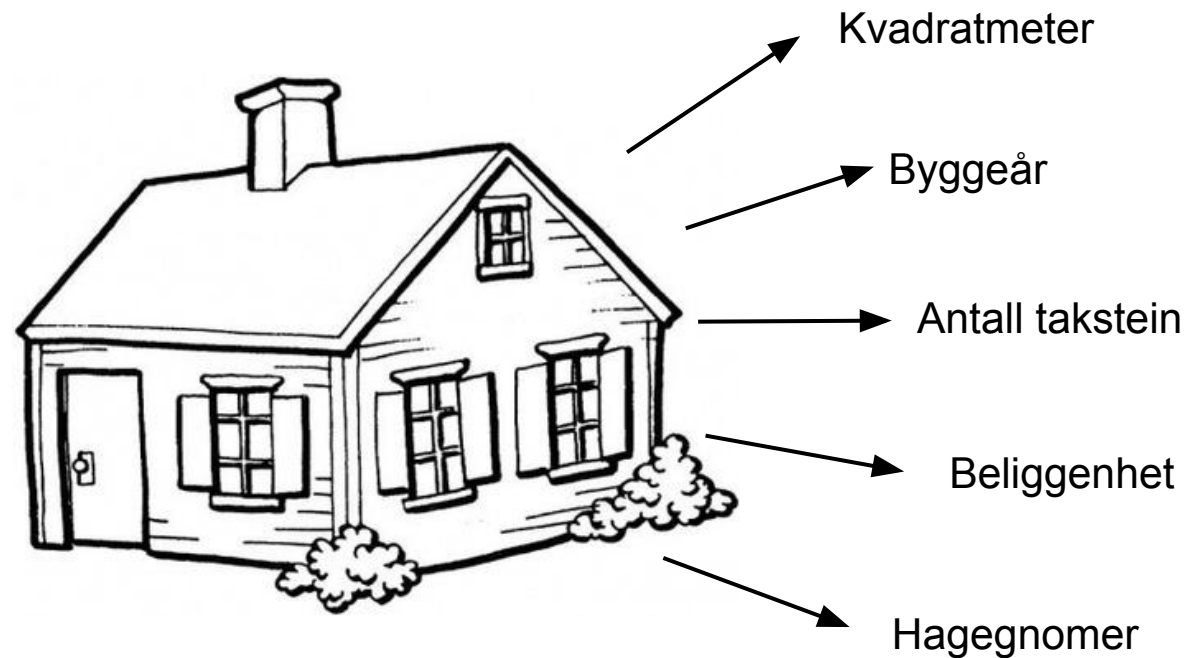
ML-algoritme

ML-model

Resultat

Test

Feature extraction



Datasett

Preprossesering

Splitte i test og
trening

Trening

ML-algoritme

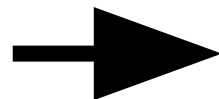
ML-model

Resultat

Test

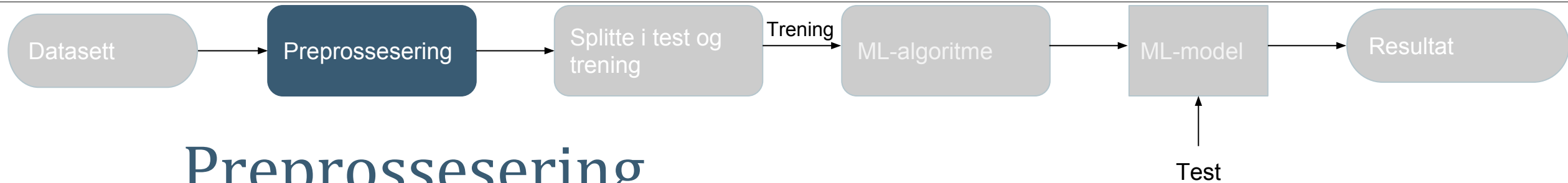
Feature selection

- Kvadratmeter
- Byggeår
- Beliggenhet
- Hagegnomer
- Antall takstein



- Kvadratmeter
- Byggeår
- Beliggenhet

Eksempel nr	Feature 1 (m ²)	Feature 2 (byggeår)	Feature 3 (beliggenhet)	Resultat (over 3M)
1	100	1960	Kjelsås	False
2	120	2001	Majorstuen	True
3	90	1970	Smestad	False



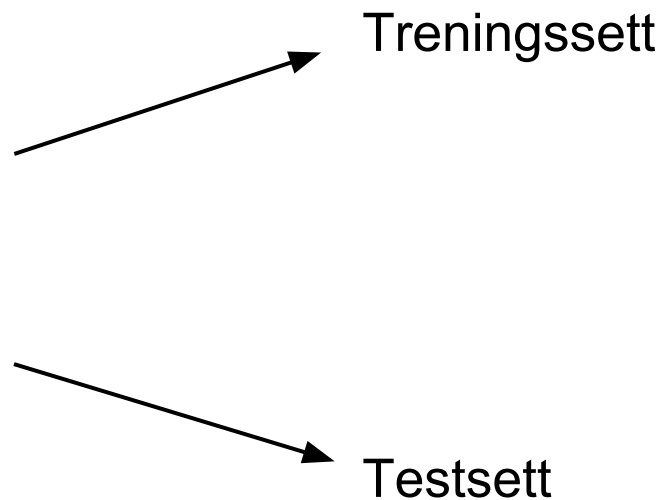
Preprossesering

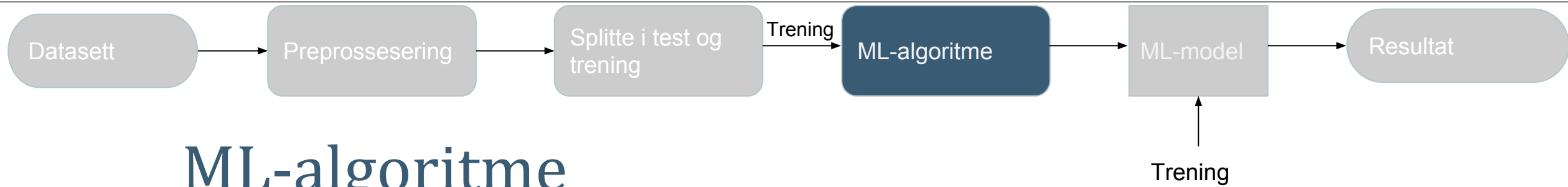
- Formaterer
 - *Er datasettet i ønsket format? (matrise/tabell)*
- Rense
 - *Mangler det features i enkelte rader?*
- Skalere
 - *Har de forskjellige featurene ulik skala?*
- Balansere
 - *Er det få eksempler av én klasse?*



Splitte i test og trening

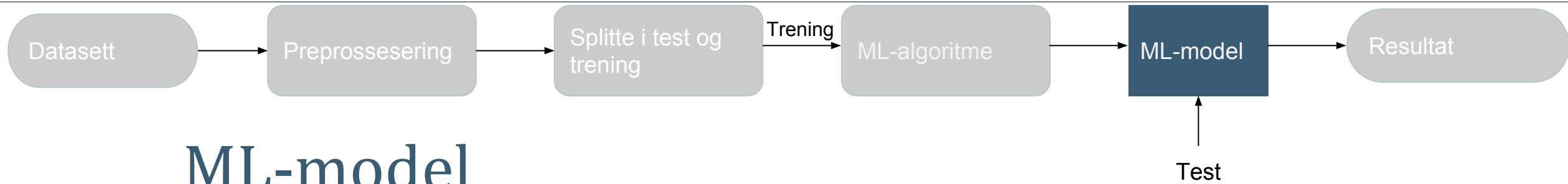
100	1966	True
95	2010	False
202	1952	False
170	1999	True
50	1986	False





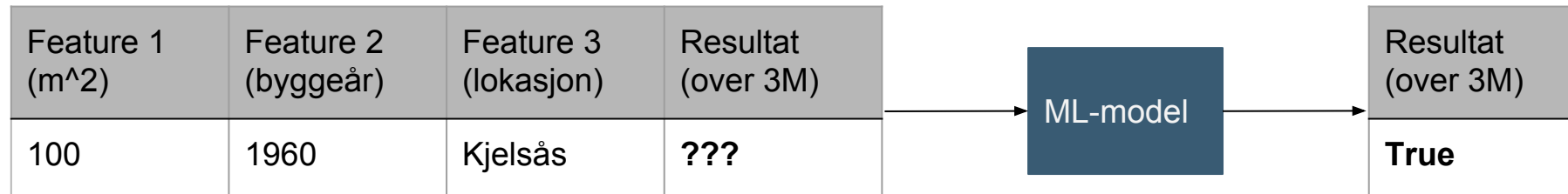
ML-algoritme

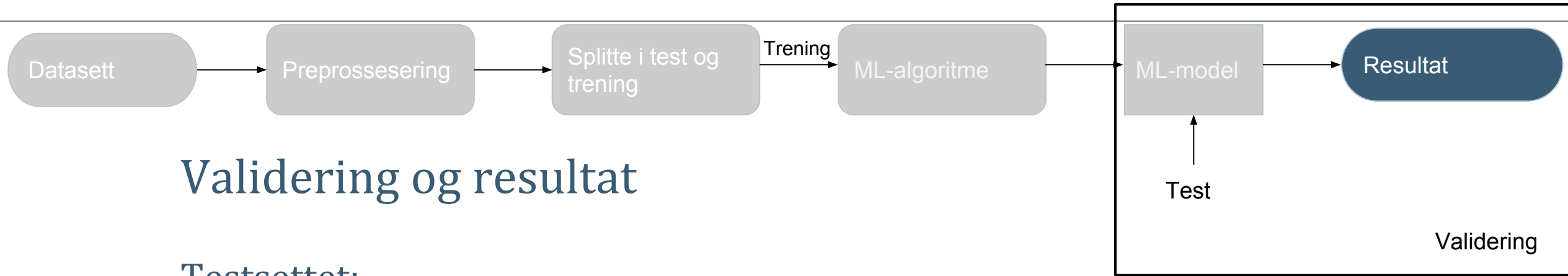
- Vil man sjeldent lage selv
 - kNN, SVM, NN, AdaBoost ...
- Har gjerne en overordnet forståelse for dem
- Må gjerne tunes gjennom parametre
- Prøver ofte flere
- Er lette å bruke



ML-model

- Selve resultatet
- Tar uklassifiserte eksempler som input og klassifiserer dem

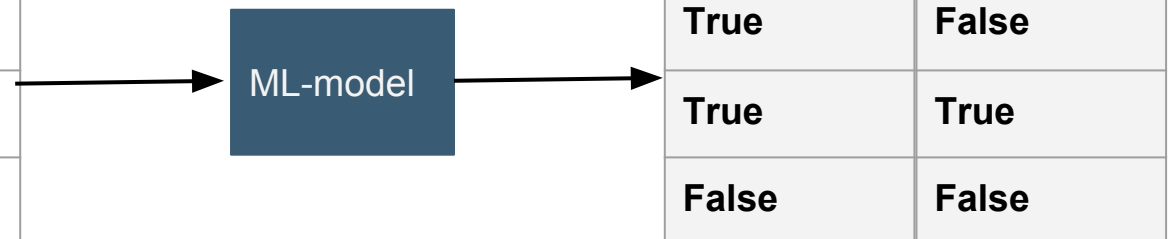


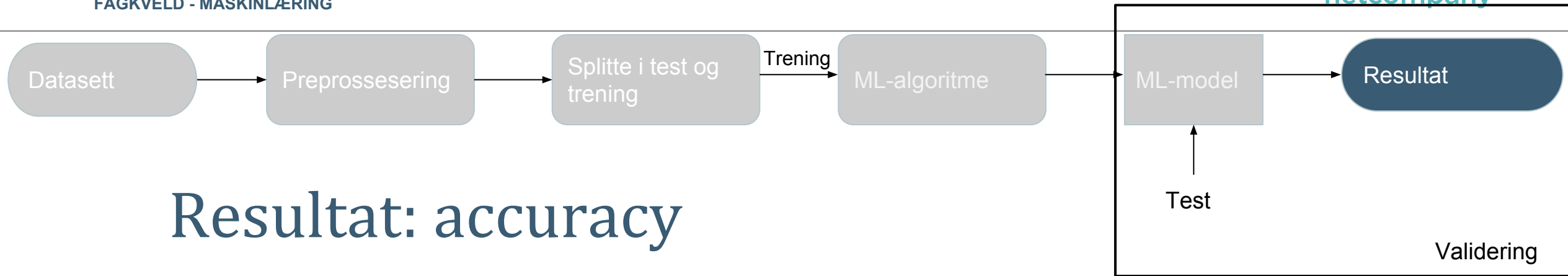


Validering og resultat

Testsettet:

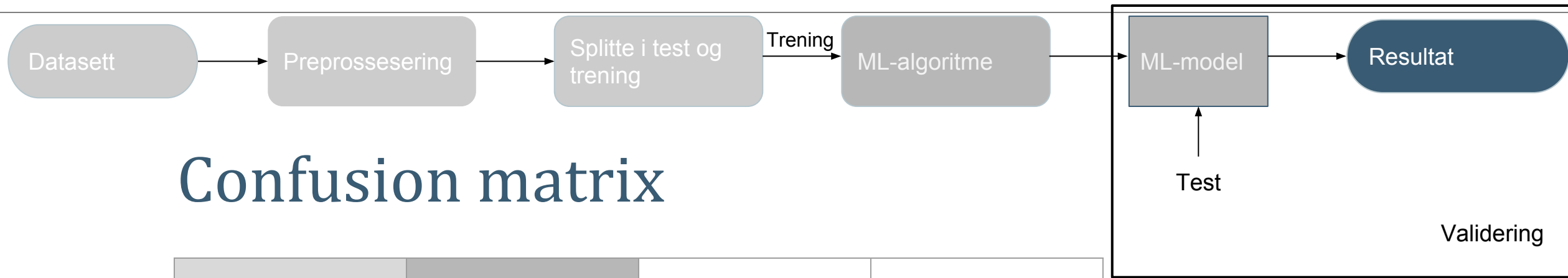
Feature 1 (m ²)	Feature 2 (byggear)	Feature 3 (lokasjon)	Resultat (over 3M)
100	1960	Kjelsås	---
120	2001	Majorstuen	---
90	1970	Smestad	---





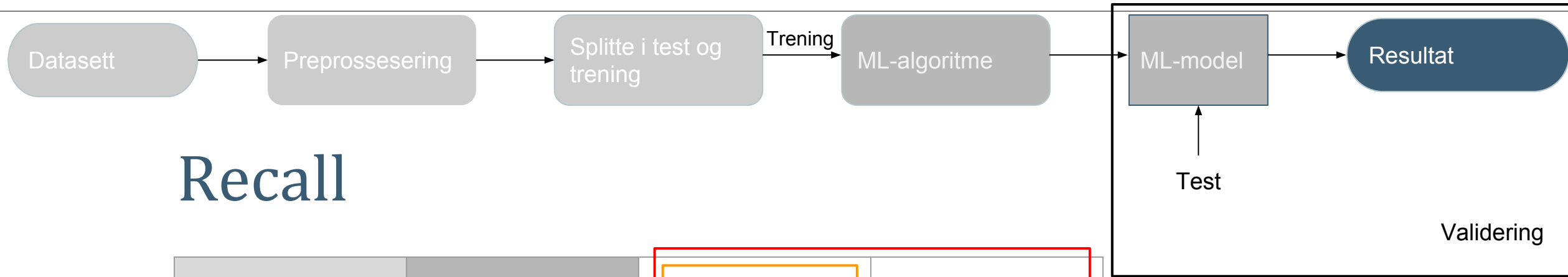
Resultat: accuracy

$$accuracy = \frac{\#riktige\ prediksjoner}{\#prediksjoner}$$



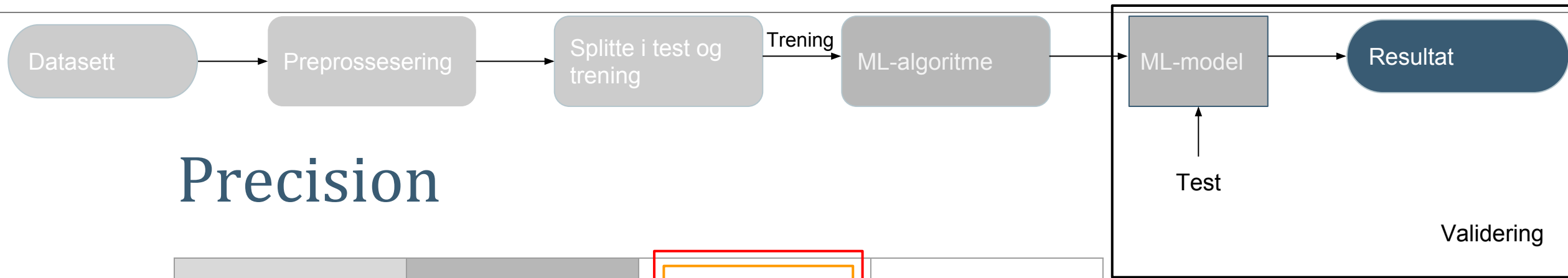
Confusion matrix

Sann klasse	Klasse 1	30	5
	Klasse 2	3	33
		Klasse 1	Klasse 2
	Predikert klasse		



Recall

Sann klasse	Klasse 1	30	5
	Klasse 2	3	33
		Klasse 1	Klasse 2
	Predikert klasse		



Precision

Sann klasse	Klasse 1	30	5
	Klasse 2	3	33
		Klasse 1	Klasse 2
	Predikert klasse		

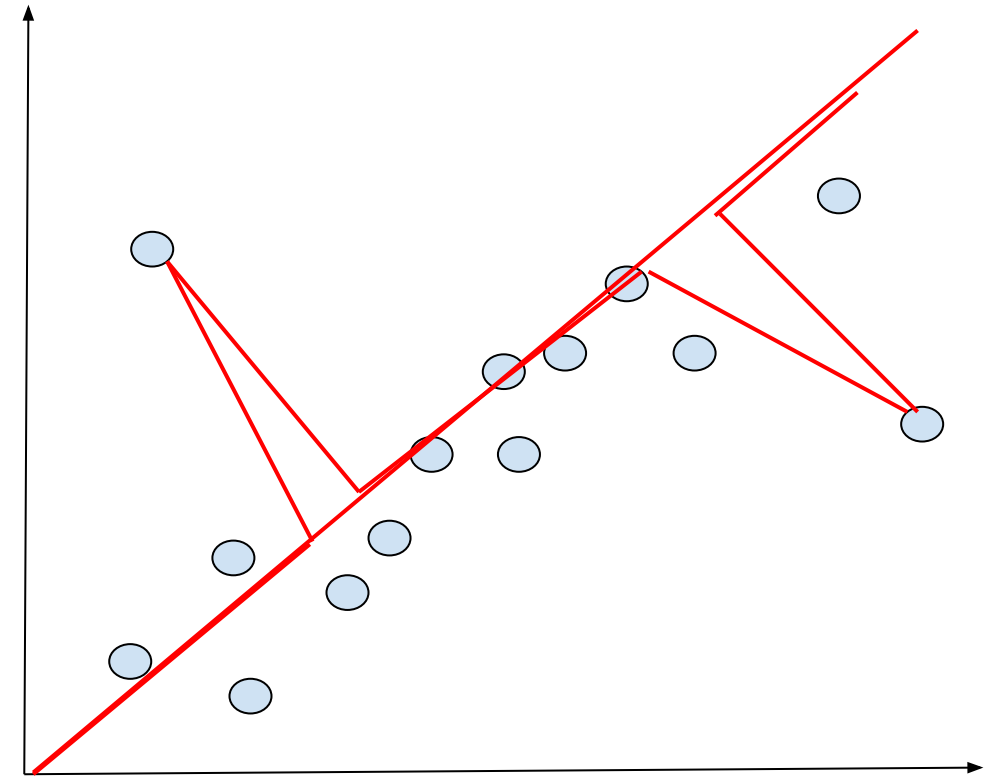


Resultat: F1-score

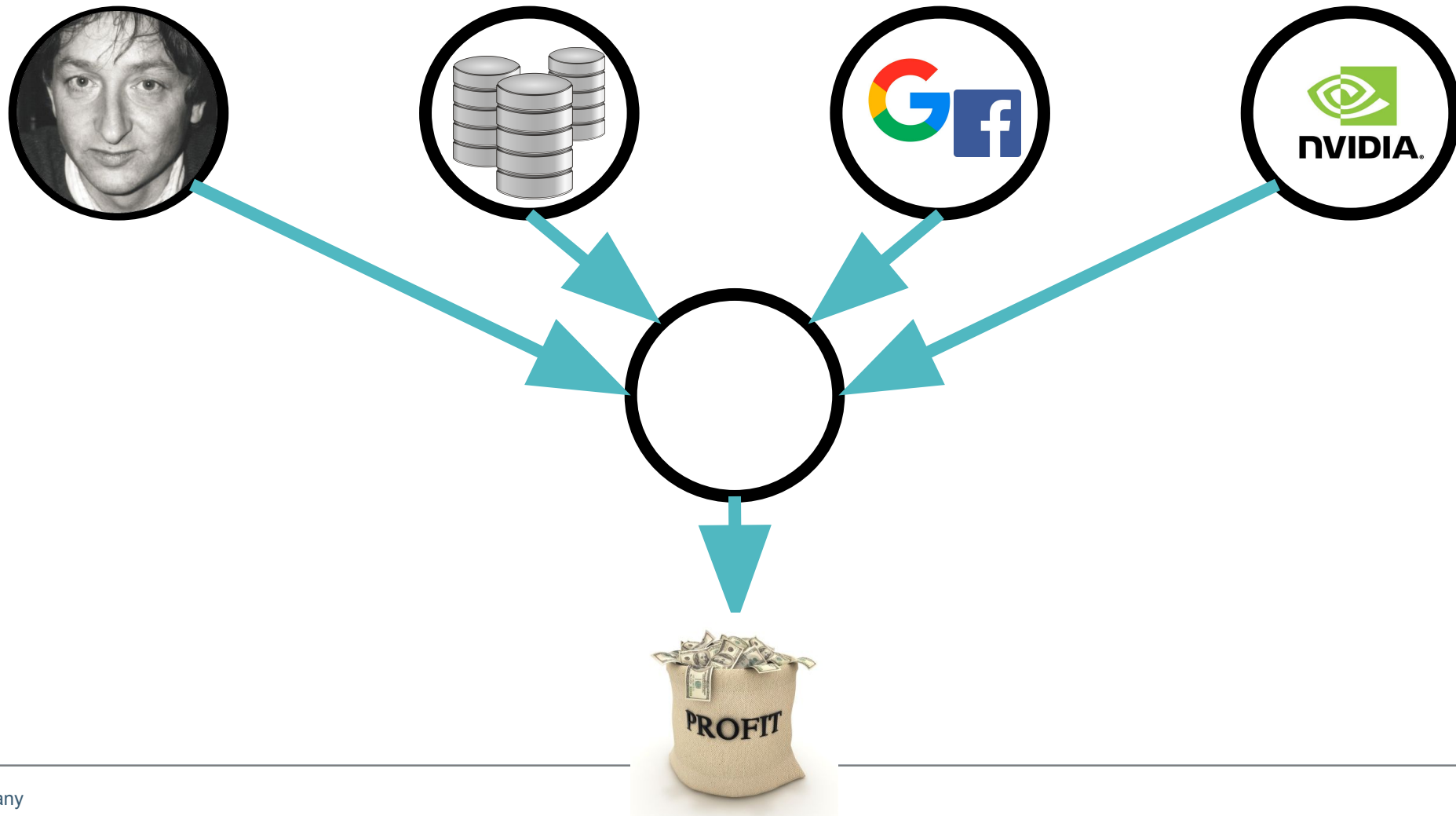
$$F_1 = 2 \times \frac{\textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}}$$

Viktig konsept: overfitting

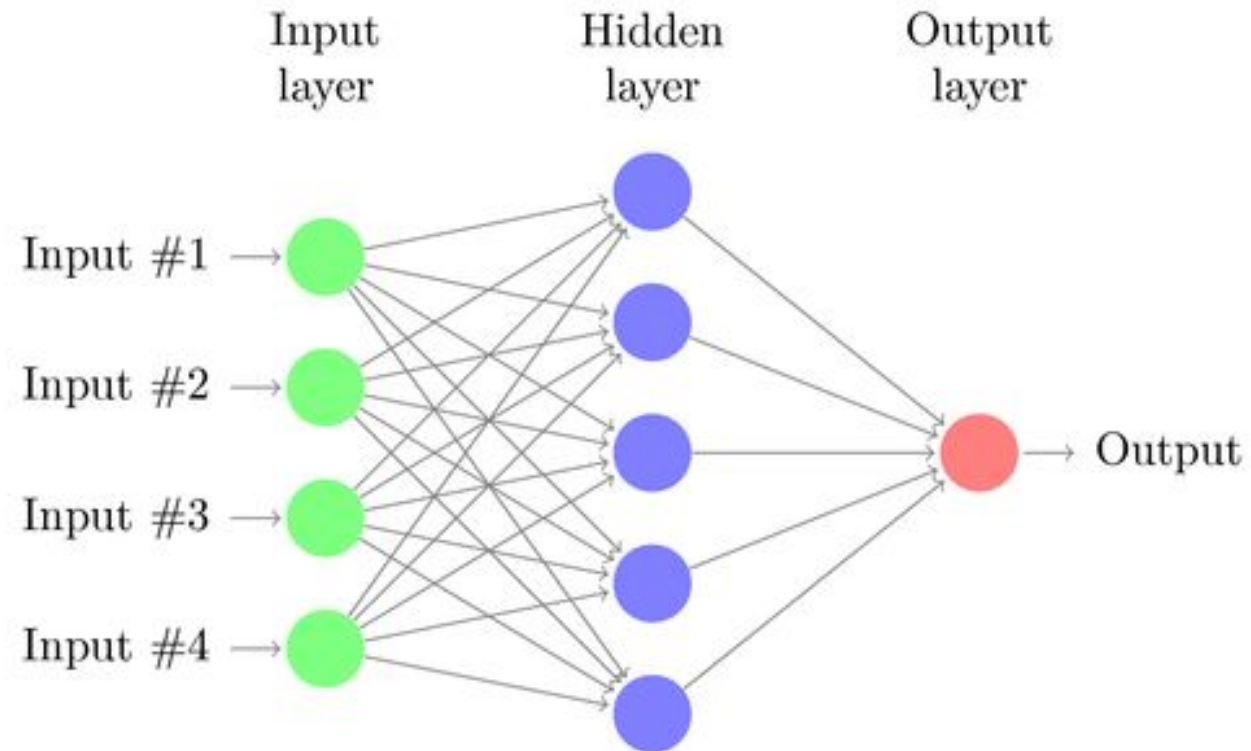
- Lærer om støyen i datasettet
- Klarer ikke å generalisere
- Lærer treningsdata *for* godt
- Er derfor vi må trene på en del av datasettet og teste på et annet



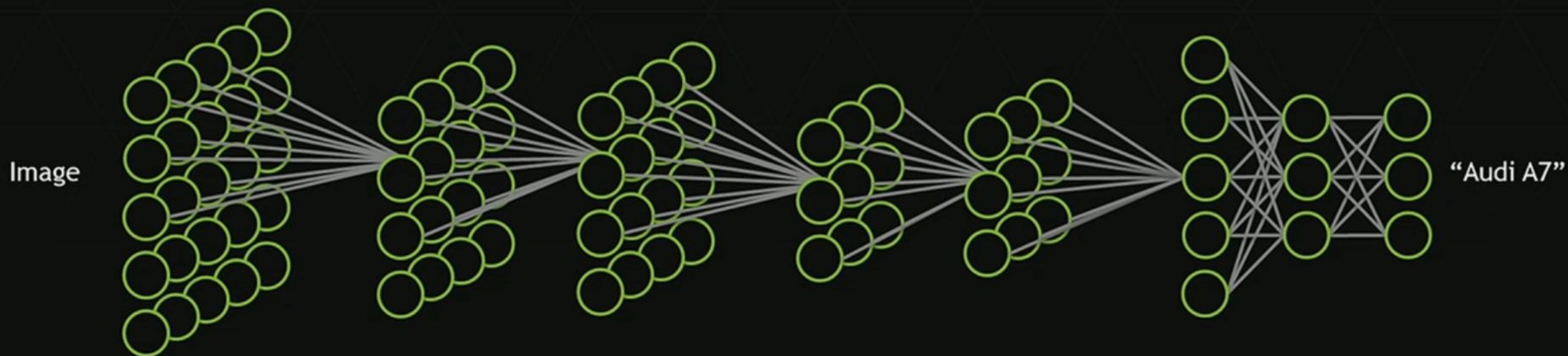
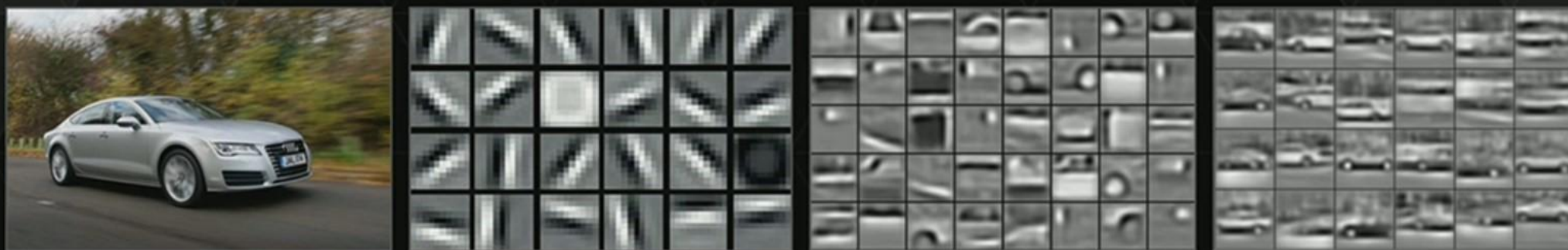
Nevrale nettverk



Nevrale nettverk



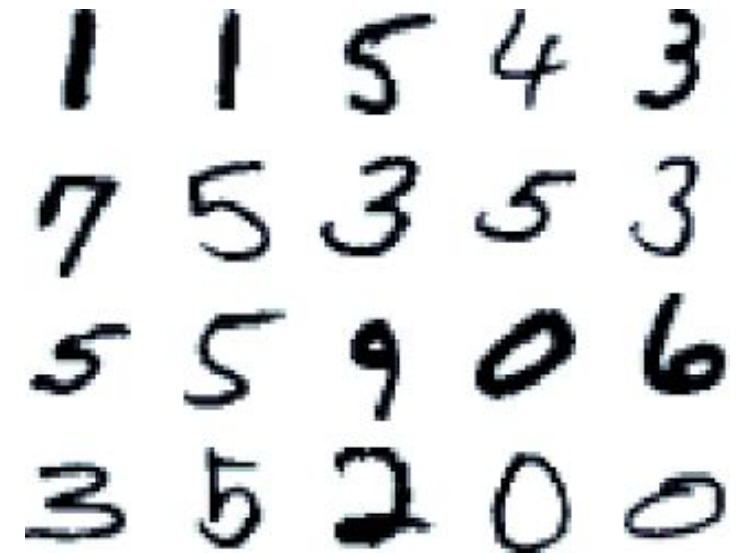
HOW A DEEP NEURAL NETWORK SEES



Oppgaver

Datasettet

- 70,000 håndskrevne tall
- Delt opp i 60,000 treningsbilder og 10,000 testbilder
- Hvert bilde er på 28x28 piksler
- Ti klasser (0 - 9)
- `data = array[antall bilder][høyde][bredde]`



Oppgaver

- Oppgave 1: Flat ut og normaliser bildene
- Oppgave 2: Bygg nettverket
- Oppgave 3: Tren og test nettverket
- Oppgave 4: Test selv
- Oppgave 5: AI app

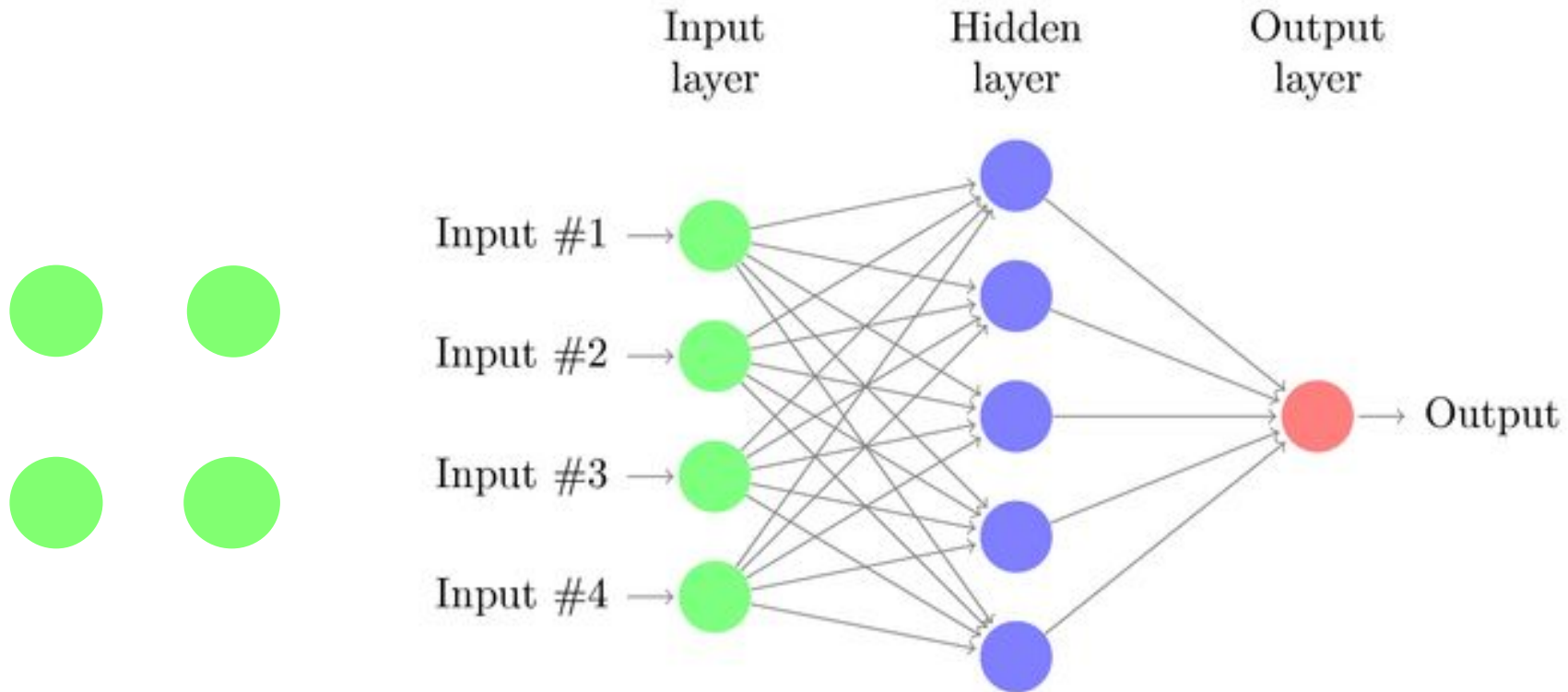


Oppgave 1 - Flat ut og normaliser bildene

1	2	3
4	5	6
7	8	9

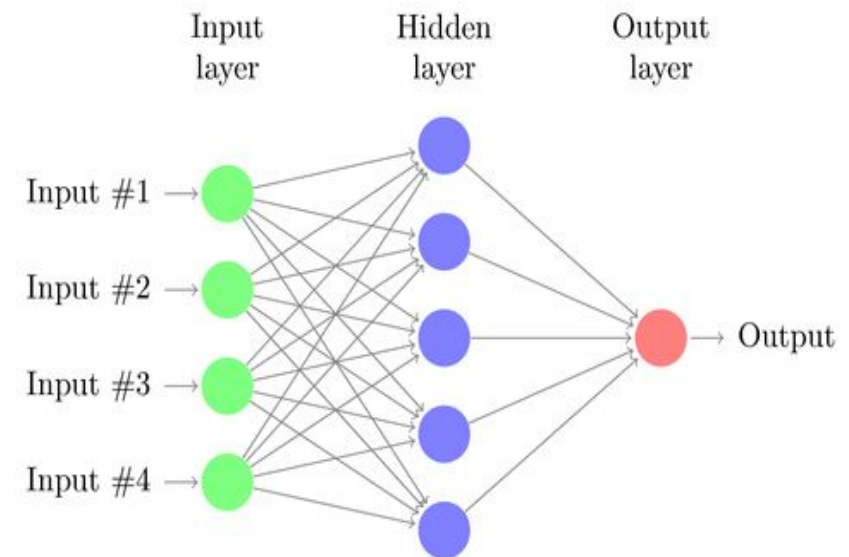
1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

Oppgave 2 - Bygg nettverket

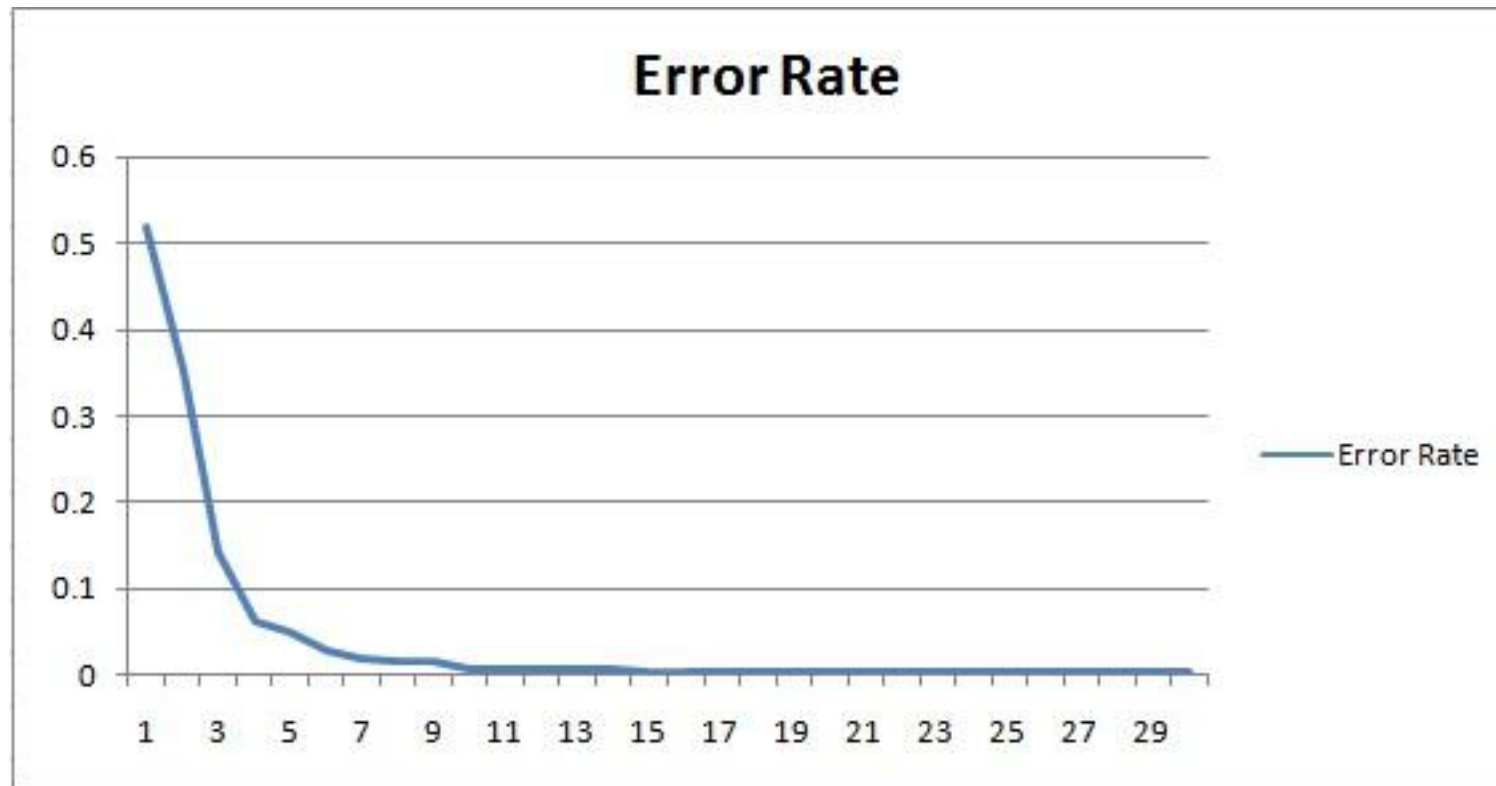


Oppgave 3 – Tren og test nettverket

1 1 5 4 3
7 5 3 5 3
5 5 9 0 6
3 5 2 0 0



Oppgave 4 – Test selv



Oppgave 5 – AI app

