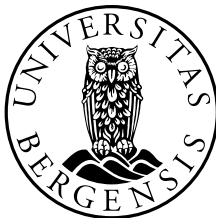


# INF161 - Innføring i Data Science

## Prosjekt

Nello Blaser



# Data forberedelse

- ▶ Lese inn data
- ▶ Kombiner spilldata fra år  $k$  med tabell fra år  $k - 1$  og lagstatistikk fra år  $k - 1$
- ▶ Feature engineering
  - ▶ Kombiner med spilldata fra år  $k - 1$
  - ▶ Kombiner med spillerdata fra år  $k - 1$  for spillere i lag fra år  $k$
- ▶ Formater data sånn at du har en skore per rad

# Modellering

- ▶ Trening-validering-test split
- ▶ Data beskrivelse (lokaliseringsmål og spredningsmål, manglende data)
- ▶ Visualiseringer
- ▶ Imputasjon
- ▶ Enkel baseline modell
- ▶ Flere modeller
- ▶ Tilpass modeller
- ▶ Velg ut modell
- ▶ Regn ut generaliserings-RMSE

# Prediksjon

- ▶ Bruk beste modell for prediksjoner for 2020
- ▶ Formater som kaggle input
- ▶ Test på kaggle (frivillig)

# Implementasjon

- ▶ Enkel nettside
- ▶ Bruk beste modell til å predikere enkle 2020 kamper