

# Machine Learning - Kunskapskontroll 1

I denna kunskapskontroll kommer ni att bygga en modell för att prediktera bilpriser.

Ni ska arbeta tillsammans i grupper om 3-4 personer. Grupperna delas in på lektionen. Är du inte på lektionen så kontakta i första hand en kurskamrat och gå med i en grupp. Kontakta Terese om du behöver hjälp att hitta en grupp. På denna kunskapskontroll kan man få betygen IG / G. För att bli godkänd så skall ni visa att ni använder koncepten vi lärt oss korrekt.

Data finns i Excel-filen "car\_price\_dataset". Excel är ett verktyg som alla bör kunna då det frekvent används i arbetslivet. Behöver du lära dig mer kring Excel rekommenderas följande video: [Grundkurs i Excel](#).

I arbetet förväntas ni:

1. Ladda in data från Excel-fil.
2. Dela upp data i X och y, sedan tränings-, validerings- och testset.
3. Utforska och analysera data (EDA).
4. Bearbeta data. Detta kan innebära att till exempel hantera tomma värden och outliers, transformation av variabler eller omvandla kategorisk data till numerisk data.
5. Skapa (minst) tre modeller och träna de på träningsdatan. Använd GridSearchCV för att optimera hyperparametrar.
6. Använd valideringsdata för att hitta den bästa modellen. Använd RMSE som utvärderingsmått.
7. Använd testdata för att utvärdera den valda modellen. Använd RMSE som utvärderingsmått.

Genomför modelleringen i en Jupyter Notebook.

Motivera era val, presentera eventuella begränsningar ni har identifierat med datan och skriv en slutsats kring er valda modell

Ladda upp era lösningar i en mapp (repository) på GitHub. Det räcker med att det finns uppladdat på en gruppmedlems GitHub. Glöm inte att göra mappen "public" så att jag kan komma åt filerna. Lämna in en länk på itslearning.

Deadline: Onsdag 4 feb 2026