

Inlämningsuppgift 7 - XGBoost
Maskininlärning och Deep learning - VT 2023
Individuell inlämning
Lämna in senast 7/6 08:00

I denna ska kunder analyseras och prediktioner på vilka som lämnar (churnar) ska göras.
[XGBoost](#) skall användas för att göra prediktionerna.

Inlämningen är uppdelad i en godkänt del och en väl godkänt del.
För godkänt krävs godkänt på godkänt delen, inklusive efter komplettering.
För väl godkänt krävs att man blir godkänd på både godkänt delen och väl godkänt delen vid första inlämningen.

En notebook ska lämnas in som innehåller kod samt svaret på frågor. Tänk på att förklara bilder och svar. Döp filen till *för_efternamn_xgboost.ipynb*

Företaget telco vill ha hjälp med att analysera kunder och hitta vilka som kommer lämna företaget för att kunna jobba för att behålla de kunderna. En churnmodell ska tränas för att hjälpa med detta problem.

För både godkänt och väl godkänt ska *churn_train.csv* och *churn_test.csv* användas.
Beskrivning av datan finns i ett separat dokument.
Samma saker ska göras för godkänt som för väl godkänt, men för väl godkänt finns ett krav på prestanda.
Modellen ska tränas på *churn_train.csv* och den ska testas på *churn_test.csv*

Följande punkter och frågor behandlas.

- Analysera datan
 - Tänk på vad som analyserats i tidigare uppgifter
 - Hur ser fördelningen av target datan ut? Är den jämnt fördelad? Om inte, kan det ställa till med problem?
- Förbered datan för att kunna användas av modellen
 - Finns det data som ej borde användas? Varför i så fall?
 - Det finns en hel del textdata som måste behandlas på något vis (finns flera olika alternativ).
- Träna en XGBoost modell
- Utvärdera modellen
 - Beräkna accuracy
 - Analysera vilka attribut som har störst påverkan
 - Rita ut en confusion matrix och analysera hur modellen gör prediktioner och vilka typer av fel den gör.
- För väl godkänt ska modellen ha en accuracy som är över 78.5% på testdatan.