

Databehandling

Kursen ingår i utbildningen: Utvecklare inom AI och maskininlärning

Kursens omfattning: 25p

Undervisningsspråk: Svenska

Förkunskaper: Inga

Valbar kurs: Nej

Utbildningsnummer och omgång: YH-01462 - 2024 – 1

Beslutsdatum för kursplan: 2024-05-22

Kursens huvudsakliga innehåll

Kursen behandlar typiska tekniker för datautforskning och databehandling, även kallat *data mining*. Kursen kommer täcka datakombinering och -aggregering, hantering av saknad data, dimensionsreducering samt "datastädning". Under kursens gång kommer exempel och övningar att utföras med Pythonpaketet Pandas. Kursen kommer även behandla begrepp som KPIer, anonymisering av data (främst för GDPR), samt grundläggande datasäkerhet. Målen med kursen är att den studerande ska kunna hantera, behandla och "städa" data samt utföra datatransformationer, lära sig metoder och tekniker för datautforskning och dataanalys och lära sig grunderna i Pythonpaketet Pandas. Kursen syftar till att ge studerande de verktyg för datautforskning och databehandling som är nödvändiga för att analysera data, dra slutsatser från data samt nyttja data för maskininlärning.

Kursens mål

Den studerande ska efter genomförd kurs med godkända betyg kunna:

Kunskaper

1. Grunder och syntax för Pandas
2. Grundläggande kännedom om GDPR och säkerhet gällande dataförvaring
3. Datakombinering och -aggregering
4. Vanliga dataformat (som csv, xml, json, yaml)
5. Visualiseringstekniker med hjälp av matplotlib, seaborn och plotly
6. Explorativ dataanalys

Färdigheter

7. Städa, aggregera, extrahera metadata och statistiska värden om data
8. Hantera och identifiera saknad eller felaktig data
9. Med vägledning kunna identifiera och definiera KPIer
10. Rita grafer för explorativ dataanalys
11. Skapa enklare dashboards med hjälp av plotly

Kompetenser

12. Självständigt kunna ta fram relevanta statistiska mått, extrahera nyckelvärden ur data samt kunna identifiera och analysera trender
13. Självständigt kunna utföra en utforskande dataanalys och dra slutsatser kring data i en verksamhet

14. Kunna hantera data på ett sätt som följer regler, praxis och etablerade arbetssätt rörande säkerhet och lagar (t.ex. GDPR) samt kunna bidra till att identifiera säkerhetsbrister i en datahanteringsprocess

Läromedel

Allt material på www.ithsdistans.se och eventuellt annat material som läraren delar ut.

Former för kunskapskontroll

Kunskapskontroller görs under kursen genom laboration och projekt med muntlig redovisning.

För varje examinationsmoment har den studerande rätt att delta på en ordinarie och två omexaminationer. Examinator avgör i vilka fall omexamination kan ges i form av komplettering.

Principer för betygssättning

Betyg sätts i form av Icke godkänt (IG), Godkänt (G) eller Väl godkänt (VG).

Icke godkänt (IG)

Den studerande har fullföljt kursen men inte nått alla mål för kursen.

Godkänt (G)

Den studerande har nått samtliga mål för kursen.

Väl godkänt (VG)

Den studerande har nått samtliga mål för kursen. Den studerande kan dessutom

- 15. Med säkerhet använda Pandas för databehandling och dataanalys**
- 16. Med säkerhet ta fram nyckeltal, statistiska mått för att analysera data**
- 17. Använda effektiva metoder i Pandas för prestandaoptimering vid behov**

Examinerande moment

Innehåll	Introduktion	Deadline	Täcker upp kursmål
Laboration			1, 4, 6, 8, 12, 13, 15, 16, 17
Projekt med muntlig redovisning			2, 3, 5, 7, 9, 10, 11, 14