

Kurs PM

Databehandling – 25 p

Beskrivning av kursen/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.

Teknologier/verktyg som ska användas:

T.ex

- Visual Studio Code
- Git/GitHub
- Python
- Pandas
- Matplotlib
- Seaborn
- Plotly
- pipenv

Utbildare

Kokchun Giang – kokchun.giang@iths.se

Datum

2021-10-18 – 2021-11-18

Mer detaljer kring innehållet som behandlas i respektive studievecka finns [här](#).

Kurslitteratur:

Bloggar, handouts, kompendier och videoklipp som utbildaren tillhandahåller:

- finns samlat [här](#)

Examinationsformer:

- Laboration
- Projekt

Betygssättning:

Studerande ska få sitt kursbetyg senast 3 veckor efter det att kursen avslutats

Betygskriterier:

Kursen bedöms med betygen Icke Godkänd, Godkänd eller Väl godkänd. Grunden för betyg avgörs av hur väl studenten har uppfyllt betygskriterierna.

Studenten ska få sitt kursbetyg senast 3 veckor efter det att kursen avslutats.

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

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
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
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

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

Examinationer:

Innehåll	Introduktion	Deadline	Täcker upp betygskriterier
Laboration	2021-10-25	2021-11-05	1, 3-10, 12-13, 15-16
Projekt	2021-11-02	2021-11-18	1-17