

DABN23, Dataanalys och ekonomi: Insamling och förberedelse av data, 7,5 högskolepoäng

Data Analytics and Business Economics: Data Retrieval and Preparation, 7.5 credits
Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Institutionsstyrelsen vid Nationalekonomiska institutionen 2025-11-06. Kursplanen träder i kraft 2025-11-06 och gäller från och med vårterminen 2026.

Allmänna uppgifter

Kursen är i ämnet dataanalys och ekonomi. Kursen ingår som valbar kurs inom magisterprogrammet Data Analytics and Business Economics. Kursen kan ingå som valfri kurs i andra magister- och masterprogram inom Lunds universitet.

Undervisningsspråk: Engelska

(Undervisningen kan ske på svenska om alla registrerade studenter behärskar svenska.)

Huvudområde

Dataanalys och
ekonomi

Fördjupning

A1N, Avancerad nivå, har endast kurs/er på grundnivå som
förkunskapskrav

Kursens mål

Kunskap och förståelse

Studenten ska tillägna sig en förståelse för:

- Python- och webbaserade API:er (Application Programming Interfaces),
- arbetsflöden för webbskrapningsprojekt,
- populära Python-bibliotek för webbskrapning och deras funktionalitet,
- utforskande analys för att styra "feature engineering",

- kodning och transformation av prediktorer,
- identifiering och modellering av interaktioner,
- hantering av saknade värden och extremvärden,
- mål, metoder och fallgropar vid variabelurval.

Färdighet och förmåga

Studenten ska ha förmåga att självständigt:

- använda Python-API:er för datainsamling,
- tolka och extrahera data från komplexa HTML-sidor,
- utveckla webbcrawlers i Python,
- planera och genomföra ett arbetsflöde för feature engineering,
- konstruera prediktorer för förbättrad modellprestanda,
- identifiera och modellera interaktioner med lämpliga sökstrategier,
- systematiskt hantera saknade värden.

Värderingsförmåga och förhållningssätt

Studenten ska ha förmåga till vidare studier inom ämnet samt förmåga att med hög grad av självständighet söka och utvärdera ämnesrelevant information. Studenten ska även ha tillräckliga färdigheter för att på egen hand skriva ett empiriskt inriktat arbete där maskininlärning utgör ett väsentligt inslag.

Kursens innehåll

Första halvan av kursen fokuserar på modern datahämtning från webben. Studenterna lär sig de juridiska och etiska ramarna för webbskrapning, grunderna i HTTP och utvecklarverktyg, samt hur man bygger robusta skrapor i Python med HTML-parsning, webbplatsomfattande crawlers och automation för formulär, inloggnings- och JavaScript-drivna sidor. Vi går även igenom hur man läser vanliga dokumentformat (CSV, PDF, Word), arbetar med offentliga och odokumenterade API:er och lagrar resultat på ett tillförlitligt sätt.

Den andra halvan utvecklar ett rigoröst arbetsflöde för datapreparering och "feature engineering" i prediktiv modellering. Vi börjar med problemformulering och korrekt datasplittring och går vidare till förbehandling som transformation, skalning, korrelationsfiltrering och andra diagnostiska steg som anpassas på träningsdata och tillämpas på validerings-/testdata. Med denna grund betonar vi utforskande visualisering – av responsvariabler, prediktorer och deras relationer – för att vägleda utformning av "features" samt modelldiagnostik före och efter initiala passningar. Kärnan i modulen är praktisk "feature engineering" och urval: kodning av kategoriska prediktorer (från enkla dummyvariabler till angreppssätt för hög kardinalitet), konstruktion av numeriska prediktorer via en-till-en-transformationer (t.ex. Box-Cox/Yeo-Johnson), basutvidgningar och "splines", många-till-många-projektioner, identifiering och skapande av interaktionstermer samt robusta angreppssätt för saknade värden och extremvärden. Vi avslutar med en samlande översikt av "feature

selection" och dess huvudklasser (intrinsic, filter, wrapper) med betoning på att undvika överanpassning.

Kursens genomförande

1. Undervisningsform: Undervisningen består av föreläsningar och övningar.

Kursens examination

1. Examinationsuppgifter: Examinationen består av en skriftlig tentamen, aktivt deltagande i föreläsningar och övningar samt inlämningsuppgifter och andra uppgifter under kursens gång. Poäng från inlämningsuppgifterna får tillgodoräknas vid examinationstillfällen under innevarande termin. Andra examinationsformer kan användas i begränsad utsträckning.

2. Begränsning av antal examinationstillfällen: –

Lunds universitet ser mycket allvarigt på fusk och kommer att vidta disciplinåtgärder mot alla slags försök till fusk i samband med tentamina eller andra examinationsformer. Plagiering betraktas som ett mycket allvarligt akademiskt brott. Det straff som universitetet kan utdela för detta, och för andra slags fusk i samband med olika former av examination, inkluderar avstängning från universitetet under en viss tidsperiod.

Om så krävs för att en student med varaktig funktionsnedsättning ska ges ett likvärdigt examinationsalternativ jämfört med en student utan funktionsnedsättning, så kan examinator efter samråd med universitetets avdelning för pedagogiskt studentstöd fatta beslut om alternativ examinationsform för berörd student.

Betyg

Betygsskalan omfattar betygsgraderna: U=Otillräckligt, E=Tillräckligt, D=Tillfredsställande, C=Bra, B=Mycket bra, A=Utmärkt

1. Betygsskala: Betyg (Benämning), Poäng alternativt procent av maxpoäng, Karakteristik

A (Utmärkt) Ett framstående resultat som är utmärkt vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet.

B (Mycket bra) Ett mycket bra resultat som karakteriseras av mycket bra teoretiskt djup, praktisk relevans, analytisk förmåga samt självständighet.

C (Bra) Ett bra resultat som karakteriseras av bra teoretiskt djup, praktisk relevans, analytisk förmåga samt självständighet.

D (Tillfredsställande) Ett resultat som är tillfredsställande vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet.

E (Tillräckligt) Ett resultat som möter minimikraven enligt ovanstående, men inte mer.

U (Underkänd) Ett resultat som är otillräckligt.

För att få godkänt på kursen behöver studenten få betyget E eller högre.

2. Sammanvägning av betyg: –

3. Betygsskalor för olika delar av kursen: –

Förkunskapskrav

Studenter som har antagits till magisterprogrammet Data Analytics and Business Economics är behöriga att läsa denna kurs. Studenter som har antagits till masterprogrammet i nationalekonomi (Economics) och som har klarat 30 högskolepoäng nationalekonomi på avancerad nivå inklusive Avancerad Ekonometri är behöriga att läsa denna kurs. För andra studenter krävs avlagd kandidatexamen med minst 30 högskolepoäng statistik varav 7,5 högskolepoäng ekonometri eller

regressionsanalys, eller avlagd kandidatexamen i företags- eller nationalekonomi med minst 15 högskolepoäng statistik varav 7,5 högskolepoäng ekonometri eller regressionsanalys.

Övrigt

1. Övergångsregler: –
2. Begränsning i giltighetstid: –
3. Begränsningar: –
4. Liknande kurser: –
5. Begränsning av förnyad examination: –