



Kursplan

Fakulteten för teknik

Institutionen för datavetenskap och medieteknik

4DV510 Datautvinning, 5 högskolepoäng

Data Mining, 5 credits

Huvudområde

Datavetenskap

Ämnesgrupp

Informatik/Data- och systemvetenskap

Nivå

Avancerad nivå

Fördjupning

A1F

Fastställande

Fastställd av Fakulteten för teknik 2019-10-28

Kursplanen gäller från och med höstterminen 2020

Förkunskaper

- Kurser inom huvudområdet datavetenskap omfattande 90 hp (inklusive examensarbete på kandidatnivå).
- 5 hp maskinläring på avancerad nivå (t.ex. 4DV660 eller motsvarande)
- Engelska 6/Engelska B eller motsvarande.

Mål

Efter slutförd kurs skall studenten kunna:

Kunskap och förståelse

- A.1 Förklara grundläggande koncept och principer för datautvinning, t.ex. avståndsmått och klustering, samt
- A.2 redogöra för de senaste rönen inom datautvinning, t.ex. tekniker och tillämpningar.

Färdighet och förmåga

- B.1 Givet ett problem och en datamängd eller dataström, planera och strukturera en datautvinningspipeline med avseende på vilka metoder som skall användas för att t.ex. dimensionsreducering, avståndsmått och klustering, samt vilken kvalitet utdata kommer att ha,
- B.2 implementera några av de enklare algoritmerna, t.ex. MDS och DBSCAN-klustering på ett effektivt sätt, samt
- B.3 använda lämpliga befintliga verktyg och programbibliotek för att utvinna data ur en given (ostrukturerad) datamängd eller dataström.

Värderingsförmåga och förhållningssätt

- C.1 Reflektera över en datautvinningstillämpning ur ett samhällsperspektiv med avseende på t ex. etiska frågeställningar kontra nytta, samt
- C.2 resonera kring datakvalitet med avseende på hur den påverkar hur utvunnen data kan användas samt hur datakvaliteten hos indata kan förbättras.

Innehåll

Kursen ger en introduktion till datautvinning och vanliga användningsområden, t ex. sökmotorer, rekommendationssystem och textutvinning. Följande moment behandlas:

- Introduktion till datautvinning
- Relationen mellan datautvinning och maskininlärning
- Hur hittar man liknande saker, t.ex. dokument och bilder?
- Hur utvinns data ur dataströmmar?
- Analys av länkar, t.ex. PageRank och HITS
- Klustingsalgoritmer, t.ex EM och Affinity Propagation
- Dimensionsreduktion, t.ex. t-SNE och Self-Organizing Maps
- Utvinning av information från text
- Tillämpningar inom webben, t.ex. rekommendationssystem och analys av sociala nätverk
- Datautvinning ur ett samhällsperspektiv, t.ex. etiska frågeställningar, affärsnytta och nya möjligheter inom t.ex. hälsa
- Datakvalitet
- Verktyg och programbibliotek för datautvinning.

Undervisningsformer

Undervisningen består av traditionella föreläsningar där teori introduceras, seminarier där tillämpning av olika metoder diskuteras utifrån ett problem samt lärarledda laborationer där praktiska färdigheter övas. Kursen innehåller även gästföreläsningar där industrirepresentanter och forskare presenterar hur och varför de använder datautvinning samt vilka metoder och verktyg de använder.

Examination

Examinationen av kursen delas in i följande moment:

Kod	Benämning	Betyg	Poäng
2001	Programmeringsprojekt	AF-skalan	2,00
2002	Muntlig tentamen	AF-skalan	1,00
2003	Programmeringsuppgifter	AF-skalan	1,00
2004	Reflektionsrapport	AF-skalan	1,00

Kursen bedöms med betygen A, B, C, D, E, Fx eller F.

Betyget A utgör det högsta betygssteget, resterande betyg följer i fallande ordning där betyget E utgör det lägsta betygssteget för att vara godkänd. Betyget F innebär att studentens prestationer bedömts som underkända.

Bedömning av de studerandes prestationer sker genom programmeringsuppgifter, muntlig tentamen, programmeringsuppgifter och en reflektionsrapport. Förnyad examination ges i enlighet med Lokala regler för kurs och examination på grundnivå och avancerad nivå vid Linnéuniversitetet.

För godkänt betyg på kursen krävs minst betyg E på samtliga moment. Slutbetyget bestäms från: Programmeringsprojekt (40%), Muntlig tentamen (20%), Programmeringsuppgifter (20%) and Reflektionsrapport (20%).

Om universitetet beslutat att en student har rätt till särskilt pedagogiskt stöd på grund av funktionsnedsättning, har examinator rätt att ge ett anpassat prov eller att studenten genomför provet på ett alternativt sätt.

Måluppfyllelse

Examinationsmomenten kopplas till lärandemålen enligt följande:

Mål	2001	2002	2003	2004
A.1		✓		
A.2		✓		✓
B.1	✓	✓		
B.2	✓		✓	
B.3	✓		✓	
C.1				✓
C.2		✓		

Kursvärdering

Under kursens genomförande eller i nära anslutning till kursen genomförs kursvärdering. Resultat och analys av genomförd kursvärdering ska skyndsamt återkopplas till de studenter som genomfört kursen. Studenter som deltar vid nästa kurstillfälle erhåller återkoppling vid kursstart. Kursvärdering genomförs anonymt.

Övrigt

Betygskriterier för A-F-skalan kommuniceras till studenten via särskilt dokument. Studenten informeras om kursens betygskriterier senast i samband med kursstart.

Kursen genomförs på ett sådant sätt att kursdeltagarnas erfarenheter och kunskap görs synlig och utvecklas. Det innebär till exempel att vi har ett inkluderande förhållningssätt och strävar efter att ingen ska känna sig exkluderad. Detta kan yttra sig på olika sätt i en kurs, till exempel genom att läraren använder sig utav könsneutrala exempel.

Kurslitteratur och övriga läromedel

Obligatorisk litteratur:

- Leskovec, Jure, Rajaraman, Anand, and Ullman, Jeffrey D., *Mining of Massive Datasets*, Cambridge University Press, senaste upplagan. Antal sidor: 400 av 511.
- Kompendium med vetenskapliga artiklar