



## Kursplan

Fakulteten för teknik

Institutionen för datavetenskap och medieteknik

1DV515 Datautvinning i praktiken, 7,5 högskolepoäng  
Data Mining in Practice

### Huvudområde

Datavetenskap

### Ämnesgrupp

Informatik/Data- och systemvetenskap

### Nivå

Grundnivå

### Fördjupning

G1N

### Fastställande

Fastställd av Fakulteten för teknik 2022-01-10

Kursplanen gäller från och med höstterminen 2022

### Förkunskaper

Grundläggande behörighet.

## Mål

### *Kunskap och förståelse*

Efter slutförd kurs skall studenten kunna

- A.1 beskriv grundläggande begrepp och principer för datautvinning, t.ex. dataförbehandling, distansmätningar och klustring
- A.2 förklara de vanligaste/typiska datautvinning algoritmerna och teknikerna för att extrahera insikter från data
- A.3 identifiera vanliga olika kategorier av datamängder baserat på datainnehållet.

### *Färdighet och förmåga*

Efter slutförd kurs skall studenten kunna

- B.1 givet en rå datamängd: förbehandla, tvätta och organisera data för att förbereda den för en datautvinningspipeline
- B.2 givet ett datautvinningrelaterat problem: (a) Tillämpa de vanligaste/typiska datautvinning algoritmerna och teknikerna för att extrahera insikter från data; och (b) Presentera resultaten från datautvinning algoritmerna på lämpligt sätt med hjälp

- av grundläggande visualiseringskoncept
- B.3 använd lämpliga (programmerings- eller icke-programmeringsbaserade) verktyg för att implementera grundläggande datautvinningspipeline.

### *Värderingsförmåga och förhållningssätt*

Efter slutförd kurs skall studenten kunna

- C.1 kritiskt reflektera över datautvinning ur ett samhällsperspektiv, med avseende på etik och värde, samt överväga etablerad kunskapsforskning och toppmodern teknik
- C.2 ta välgrundade beslut för att identifiera vilka datautvinningsuppgifter och algoritmer/tekniker som kan/bör tillämpas med tanke på olika krav för datautvinning och olika kategorier av datamängder.

## Innehåll

Kursen introducerar grundläggande datautvinning på ett tillämpat sätt genom att fokusera på begränsade verkliga problem (med kända omfattningar) och grundläggande principer som kan tillämpas direkt med eller utan programmeringskunskap. Några exempel på tillämpningsområden är (men inte begränsat till): samhällsvetenskap (digitala humaniora, inlärningsanalys, etc.), sjukvård och biovetenskap (beräkningsbiologi, digital hälsa, etc.), skogsbruk och träteknik och andra områden där stora datamängder är tillgängliga och som behöver undersökas. Några exempel på algoritmer och tekniker som omfattas av kursen är: sökmotorer, nätverksanalys och community detection, klustring, text utvinning och grundläggande visualisering. Följande är en sammanfattning av de ämnen som behandlas av kursen:

- Introduktion till datautvinning.
- Förbehandling av data: rensning, filtrering och strukturering.
- Hur man hittar liknande saker, t.ex. dokument och bilder.
- Nätverksanalys, t.ex. community detection och centrality measures.
- Klustringsalgoritmer, t.ex. K-means och Hierarchical Clustering.
- Dimensionality Reduction, t.ex. PCA och MDS.
- Extrahering av information från ostrukturerad text.
- Datautvinning ur ett samhällsperspektiv med avseende på t.ex. etiska frågor, affärsvärde och hälsa.
- Verktyg och programvarubibliotek för datautvinning.

## Undervisningsformer

Undervisningen består av föreläsningar, tillämpad handledning och lärarledda laborationer. Kursen kan också innehålla gästföreläsningar där representanter från den akademiska världen och industrin diskuterar hur och varför de använder datautvinning i sitt arbete.

## Examination

Examinationen av kursen delas in i följande moment:

Kod	Benämning	Betyg	Poäng
2201	Uppgifter	AF-skalan	2,00
2202	Programmeringsprojekt	AF-skalan	2,50
2203	Muntlig Tentamen	AF-skalan	2,00
2204	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 individuella programmeringsuppgifter, muntlig tentamen, ett programmeringsprojekt 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%), Programmeringsuppgifter (30%), Muntlig tentamen (20%), and Reflektionsrapport (10%).

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	2201	2202	2203	2204
A.1	<input checked="" type="checkbox"/>			
A.2	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	
A.3	<input checked="" type="checkbox"/>			
B.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
B.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
B.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
C.1			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
C.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

### 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: 100 av 511.

James, Gareth, Witten, Daniela, Hastie, Trevor, & Tibshirani, Robert. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: Springer. Antal sidor: 50 av 426

Zaki, Mohammed J., & Meira Jr., Wagner. (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press. Antal sidor: 50 av 604

Kompendium med vetenskapliga artiklar. Antal sidor: 60