

**Kursplan för:**

## **Datateknik AV, Datamining, 6 hp**

Computer Engineering MA, Data Mining, 6 credits

### **Allmänna data om kursen**

|                               |   |
|-------------------------------|---|
| <b>Kurskod</b>                | DT047A  |
| <b>Ämne/huvudområde</b>       | Datateknik  |
| <b>Nivå</b>                   | Avancerad   |
| <b>Inriktning (namn)</b>      | Datamining  |
| <b>Högskolepoäng</b>          | 6.0   |
| <b>Fördjupning vs. Examen</b> | A1N , Kursen ligger på avancerad nivå och har endast kurs(er) på grundnivå som förkunskapskrav. |
| <b>Utbildningsområde</b>      | Teknik 100%   |
| <b>Ansvarig institution</b>   | Data- och elektroteknik   |
| <b>Inrättad</b>               | 2018-10-01  |
| <b>Fastställd</b>             | 2018-10-11  |
| <b>Senast reviderad</b>       | 2022-11-24  |
| <b>Giltig fr.o.m</b>          | 2023-01-01  |

### **Syfte**

Studenten ska utveckla en grundläggande förståelse för aktuella maskininlärningsmetoder för informationsutvinning (datamining) ur stora mängder data. Studenten ska utveckla färdigheter i att hitta intressanta mönster och bygga prediktionsmodeller genom explorativ dataanalys med hjälp av dataanalysverktyg baserade på R eller Python och kunna förbereda data, tolka resultat och kritiskt utvärdera resultat. Studenten ska visa förmåga att tillämpa kunskaperna i ett verkligt projekt inom områden såsom big data-visualisering, affärsdata för omvärldsanalys, beslutsstödssystem, dataanalys av text, webbsidor, sensor-/positionsdata, kontextmedvetna applikationer, intelligenta agenter eller kognitiv radio.

## Lärandemål

Studenten ska kunna:

- Diskutera vilka verkliga tillämpningar av datamining som är realistiska och etiska
- Utvinna information med hjälp av ett verktyg som till exempel skriptspråket R, Python-biblioteken Orange eller SciKit Learn eller egna implementationer av algoritmer
- Förbereda data, tolka data och utvärdera resultat
- Identifiera variabler av betydelse i en multivariat datauppsättning
- Utvinna associationsregler och utvärdera deras reliabilitet
- Utveckla och validera prediktionsmodeller
- Följa en standardiserad arbetsprocess för tillförlitlig problemanalys, modellering och utvärdering
- Tillämpa datamining-tekniker på ett litet problem från verkliga livet

## Innehåll

- Användningsområden för data mining
- Data och kunskapsrepresentation (relationer, attribut, glesa matriser, tabeller, beslutsträd, regler)
- Bayesiansk statistik
- Associativa och sekventiella mönster
- Grundläggande algoritmer
- Dataklustring
- Datakategorisering
- Datarengöring
- Datavisualisering
- Associationsregler
- Dataprediktering
- Laborationer på verktyget R och/eller Python
- Projekt

## Behörighet

120 hp avslutade kurser, inkluderande följande:

Datateknik GR (AB), innefattande Databaser, modellering och implementering, 6 hp och en programmeringskurs, 6 hp.

Matematik GR (A), 30 hp, innefattande Matematisk statistik, 6 hp.

## Urvalsregler

Urval sker i enlighet med Högskoleförordningen och den lokala antagningsordningen.

## Undervisning

Kursen kan komma att erbjudas som campuskurs eller som webbaserad distanskurs. Studieinsatsen uppskattas till cirka 160 timmar.

## Examination

**I101:** Projektval, 0,0 hp

**Betygsskala:** U, G

**L101:** Laboration, 0,5 hp

**Betygsskala:** U, G

**P101:** Projektpresentation, 2,0 hp

**Betygsskala:** U, G

**T101:** Tentamen, 3,5 hp

**Betygsskala:** 7-gradig betygsskala, A-F o Fx

Slutbetyget baseras på en sammanvägning av tentamens- och projektresultat.

Ämnets betygskriterier hittas på [www.miun.se/betygskriterier](http://www.miun.se/betygskriterier).

Om en student har ett besked från samordnaren vid Mittuniversitetet om pedagogiskt stöd vid funktionsnedsättning, har examinator rätt att ge anpassad examination för studenten.

Om tentamen på campus inte får genomföras enligt besked från rektor, eller den denne delegerat rätten till, gäller följande: Tentamen T101, kommer att ersättas med två delar, webbexamination och uppföljning. Inom tre veckor efter webbexaminationen kommer ett urval av studenterna att kontaktas och få svara på frågor angående genomfört prov. Uppföljningen består av frågor om genomförandet av webbexaminationen och de svar som studenten skickat in.

### **Begränsning av examination**

Studenter registrerade på denna version av kursplan har rätt att examineras 3 gånger inom loppet av 1 år enligt angivna examinationsformer. Därefter gäller examinationsform enligt senast gällande version av kursplan.

### **Betygsskala**

På kursen ges något av betygen A, B, C, D, E, Fx och F. A - E är Godkänt, Fx och F är underkänt.

## Litteratur

### Obligatorisk litteratur

**Författare/red:** Witten, Frank, Hall

**Titel:** Datamining - Pratical Machine Learning Tolls and Techinques

**Upplaga:** Third edition 2011 or later

**Förlag:** Elsvier

### Referenslitteratur

**Författare/red:** Ganguly et al

**Titel:** Knowledge discovery from sensor data

**Upplaga:** 2009 or later

# Signature page

This document has been electronically signed  
using eduSign.

eduSign