



Kursplan för:

Datateknik AV, Datamining, 7,5 hp

Computer Engineering MA, Data mining, 7.5 Credits

Allmänna data om kursen

Kurskod	DT044A
Ämne/huvudområde	Datateknik
Nivå	Avancerad
Inriktning (namn)	Datamining
Högskolepoäng	7.5
Fördjupning vs. Examen	A1N , Kursen ligger på avancerad nivå och har endast kurs(er) på grundnivå som förkunskapskrav.
Utbildningsområde	Teknik 100%
Ansvarig avdelning	Avdelningen för informations- och kommunikationssystem
Inrättad	2014-03-17
Fastställd	2014-11-24
Senast reviderad	
Giltig fr.o.m	2014-07-01

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 som R, Weka eller Orange 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 t.ex. skriptspråket R, Python-biblioteket Orange, Java-baserade verktyg Weka 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
- Laboration på verktyget R, Orange eller Weka
- Projekt

Behörighet

Tidigare studier 120 hp inkluderande:

- 30 hp matematik inklusive kurserna Matematisk statistik, 6 hp samt Matematisk modellering, 6 hp
- Java, 6 hp
- Databaser, modellering och implementering, 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 400 timmar.

Examination

0,0 hp, I101: Projektval

Betyg: Godkänd eller underkänd

0,5 hp, L101: Laboration

Betyg: Godkänd eller underkänd

3,5 hp, T101: Tentamen

Betyg: A, B, C, D, E, Fx och F

3,5 hp, P101: Projektpresentation

Betyg: Godkänd eller underkänd

Slutbetyget baseras på en sammanvägt tentamens- och projektresultat.

Betyg: A, B, C, D, E, Fx och F. A-E är godkända betyg, Fx och F underkända.

Ämnets betygskriterier hittas på www.miun.se/betygskriterier.

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 - Practical Machine Learning Tools and Techniques

Upplaga: Third edition 2011 or later

Förlag: Elsevier

Referenslitteratur

Författare/red: Ganguly et al

Titel: Knowledge discovery from sensor data

Upplaga: 2009 or later