



# INSTITUTIONEN FÖR FILOSOFI, LINGVISTIK OCH VETENSKAPSTEORI

## **LT2222 Maskininlärning för statistisk datalingvistik: inledning, 7,5 högskolepoäng**

Machine learning for statistical NLP: introduction, 7.5 credits

*Avancerad nivå / Second Cycle*

---

### **Fastställande**

Kursplanen är fastställd av Institutionen för filosofi, lingvistik och vetenskapsteori 2020-05-25 att gälla från och med 2020-08-31, höstterminen 2020.

*Utbildningsområde:* Naturvetenskapligt 100 %

*Ansvarig institution:* Institutionen för filosofi, lingvistik och vetenskapsteori

### **Inplacering**

Kan också ges som fristående kurs.

Kursen kan ingå i följande program: 1) Språkteknologi, masterprogram (H2MLT)

#### *Huvudområde*

Språkteknologi

#### *Fördjupning*

A1F, Avancerad nivå, har kurs/er på avancerad nivå som förkunskapskrav

### **Förkunskapskrav**

För tillträde till kursen krävs antingen godkänt resultat på båda kurserna:

LT2001 Introduktion till programmering, 7,5 hp

LT2002 Introduktion till formell lingvistik, 7,5 hp

eller motsvarande språkteknologisk kompetens.

### **Lärandemål**

Efter godkänd kurs ska studenten kunna:

*Kunskap och förståelse*

- redogöra för användningen av språkliga data i tillämpningar inom språkteknologi,
- visa förståelse för grundläggande metodologi inom, och likheter mellan, språkteknologi och maskininlärning,
- visa förståelse för den matematiska grunden för kvantitativa angreppssätt inom språkteknologi på en begreppslig nivå,

*Färdigheter och förmåga*

- tillämpa etablerade utvärderingsmetoder och tolka resultaten av dessa,
- tillämpa enkla statistiska språkteknologiska algoritmer och angreppssätt,
- använda verktyg för dataanalys och maskininlärning,
- använda kollaborativa verktyg för utveckling och dokumentation av mjukvara,

*Värderingsförmåga och förhållningssätt*

- jämföra och välja lämpliga angreppssätt för maskininlärning och statistik.

**Innehåll**

Kursen kommer att täcka följande områden:

- representation av data för maskininlärning och forskning om mänskligt språk,
- design av arbetsflöden (pipelines) för data,
- vetenskaplig metodologi inom maskininlärning för språkteknologi, med hänsyn till etiska och yrkesmässiga frågor,
- grunderna för maskininlärning i termer av informationsteori, sannolikhetslära, statistik och linjär algebra,
- klassiska maskininlärningsmetoder såsom Support Vector Machines (SVM) och logistisk regression,
- grunderna för perceptroner, neurala nätverk och Stochastic Gradient Descent.

**Former för undervisning**

De huvudsakliga undervisningsmetoderna är föreläsningar, demonstrationer och praktiska övningsuppgifter.

*Undervisningsspråk:* engelska

**Former för bedömning**

Kursen bedöms utifrån individuella hemuppgifter, samt eventuellt skriven tentamen och/eller quiz.

Om student som underkänts två gånger på samma examinerande moment önskar byte av examinator inför nästa examinationstillfälle, ska sådan begäran inlämnas skriftligt till institutionen och bifallas om det inte finns särskilda skäl däremot (HF 6 kap § 22).

I det fall en kurs har upphört eller genomgått större förändringar ska student garanteras minst tre examinationstillfällen (inklusive ordinarie examinationstillfälle) under en tid av minst ett år, dock som längst två år efter det att kursen upphört/förändrats. Vad avser praktik och VFU gäller motsvarande, men med begränsning till endast ett ytterligare examinationstillfälle.

Komplettering av examinerad studentprestation medges.

### **Betyg**

På kursen ges något av betygen Väl godkänd (VG), Godkänd (G) och Underkänd (U).

### **Kursvärdering**

Studenter som deltar i eller har avslutat kursen ges möjlighet att anonymt framföra erfarenheter av och synpunkter på denna i en kursvärdering. En sammanställning av kursvärderingen och kursansvarig lärares reflektion tillgängliggörs för studenterna inom rimlig tid efter kursslut. Nästa gång kursen ges presenteras sammanställningen och eventuella genomförda åtgärder.

### **Övrigt**

Kan ej ingå i examen tillsammans med LT2202 eller LT2212.