

# Text Mining

Fristående- och programkurs

6 hp

Text Mining

732A92

Gäller från: 2017 HT

**Fastställd av**

Filosofiska fakultetens kvalitetsnämnd

**Fastställandedatum**

2016-04-13

**Revideringsdatum**

2016-09-08

**Gavs sista gången**

HT 2022

**Ersätts av**

732A81

## Huvudområde

Statistik

## Utbildningsnivå

Avancerad nivå

## Fördjupningsnivå

A1N

## Kursen ges för

- Masters Programme in Statistics and Machine Learning

## Förkunskapskrav

- Kandidatexamen om 180hp (eller motsvarande) inklusive självständigt arbete 15hp på fördjupningsnivå inom något av följande ämnen:
  - statistik
  - matematik
  - tillämpad matematik
  - datavetenskap
  - teknik
- Godkända kurser i:
  - kalkyl
  - linjär algebra
  - statistik
  - programmering
- Engelska 6/B  
(Undantag för svenska)

## Lärandemål

Efter avslutad kurs skall den studerande på en avancerad nivå kunna:

- använda standardmetoder för informationssökning och informationsutvinning,
- tillämpa textbehandlingsmetoder för att förbereda dokument för statistisk modellering,
- tillämpa relevanta statistiska modeller för analys av textinformation och tolka resultat på ett korrekt sätt,
- använda statistiska modeller för prediktion av textinformation,
- utvärdera prestanda av statistiska modeller för textmaterialet.

## Kursinnehåll

Kursen behandlar hur textinformation kan utvinnas, lingvistiskt bearbetas och därefter analyseras kvantitativt med formella statistiska metoder och modeller. Kursen utnyttjar expertkunskaper från databashantering, beräkningslingvistik och statistik.

Följande ämnen är inkluderade:

Introduktion till och överblick av kvantitativ textanalys och dess tillämpningar;  
Informationsutvinning; Webbcrawling; Informationssökning; Tf-idf;  
Vektorrymsmodeller; Textbehandling; Bag of Words modellen; N-grams;  
Gleshet och utjämning för text; Dokumentklassificering; Sentimentanalys;  
Modellutvärdering; Topic modeller.

## Undervisnings- och arbetsformer

Kursen består av föreläsningar, datorövningar och ett projekt. Föreläsningarna ägnas åt genomgång av koncept och metodik. Datorövningarna ägnas åt praktisk tillämpning av Text Mining redskapen. Projektarbetet ger praktisk erfarenhet av lösning av ett specifikt Text Mining problem. Utöver detta ska den studerande utöva självstudier.

Undervisningsspråk: engelska.

## Examination

Skriftlig redovisning av projektarbetet. Skriftliga laborationsuppgifter. Detaljerad information återfinns i studiehandledningen.

Om LiU:s koordinator för studenter med funktionsnedsättning har beviljat en student rätt till anpassad examination vid salstentamen har studenten rätt till det. Om koordinatören istället har gett studenten en rekommendation om anpassad examination eller alternativ examinationsform, får examinator besluta om detta om examinator bedömer det möjligt utifrån kursens mål.

Studerande, vars examination underkänts två gånger på kursen eller del av kursen, har rätt att begära en annan examinator vid förnyat examinationstillfälle.

Den som godkänts i prov får ej delta i förnyat prov för högre betyg.

## Betygsskala

ECTS, EC

## Övrig information

Planering och genomförande av kurs ska utgå från kursplanens formuleringar. Den kursvärdering som ska ingå i varje kurs ska därför behandla frågan om hur kursen överensstämmer med kursplanen.

Kursen bedrivs på ett sådant sätt att både mäns och kvinnors erfarenhet och kunskaper synliggörs och utvecklas.

## Institution

Institutionen för datavetenskap