

# Studieplan for Analyse av Big Data (15 studiepoeng) (Høst 2015)

## Fakta om programmet

**Studiepoeng:** 15

**Studiets varighet:** ½ år

**Undervisningsspråk:** Engelsk

**Studiested:** Høgskolen i Østfold, studiested Halden

## Kontakt

Antonella Serrano

**Telefon:** +47 696 08 096

**E-post:** videre@hiof.no

### *Studieprogramansvarlig*

HiØ VIDERE

## Innholdsfortegnelse

- Informasjon om studiet
- Hva lærer du?
- Opptak
- Oppbygging og gjennomføring
- Jobb og videre studier
- Studieplanen er godkjent og revidert
- Studiemodell

## Informasjon om studiet

Dette er et betalingsstudium. For nærmere informasjon om studieavgift, vennligst kontakt HiØ VIDERE.

Målgruppen er personer som ønsker å øke sin kompetanse innen metoder for analyse av store datamengder.

Anbefalte forkunnskaper er grunnleggende kompetanse i:

- programmering,
- lineær algebra,
- sannsynlighetsteori,
- statistikk,
- kalkulus og
- optimalisering.

## Hva lærer du?

### Studiets læringsutbytte

#### **Kunnskaper:**

Studenten

- har oversikt over alle de viktigste metodene innen dataanalyse.
- har kunnskap om de utfordringer som i praksis finnes innen dataanalyse, for eksempel overfitting, manglende verdier og klassifiseringer som har forskjellig kostnad.
- har fått bedre kunnskaper i grunnleggende emner som numerisk optimering og statistiske metoder relatert til dataanalyse, for eksempel Bayesianske metoder brukt av søkemotorer som Google og i spam filter.

#### **Ferdigheter:**

Studenten kan

- tilrettelegge eller kode om data slik at det passer for dataanalyse og maskinlæringsalgoritmer.
- velge rett verktøy for en gitt type data.
- avgjøre hvor gode resultater man faktisk har ved hjelp av enkel statistisk analyse av for eksempel klassifiseringsnøyaktighet.
- bruke dataanalyse i praktiske applikasjoner.

#### **Generell kompetanse:**

- Studenten har kompetanse til å behandle og analysere data av vilkårlig type.

## Opptak

Generell studiekompetanse eller realkompetanse.

## Oppbygging og gjennomføring

### Studiets oppbygging og innhold

Studiet består av ett emne med et omfang på 15 studiepoeng.

Dette emnet gir en bred introduksjon i dataanalyse og statistisk mønster-gjenkjenning. Du blir kjent med klassiske, så vel som moderne dataanalysemetoder. I tillegg vil du lære om metodenes fordeler og ulemper, samt om når og hvordan du skal benytte dem.

Emnet tar for seg følgende temaer:

Del 1: Course Introduction, linear regression and logistic regression

Del 2: Overfitting problem, statistical view and probability density estimation

Del 3: Decision tree, Random forest, Bayesian Network and Hidden Markov Model

Del 4: Data pre-processing, generalization and advanced models

## Organisering og læringsformer

Studiet går over et semester med fire 2-dagers samlinger på dagtid, totalt 8 dager. Det gis veiledning via e-post mellom samlingene.

**Arbeidskrav:** Minimum 75% deltakelse på undervisningssamlinger. Tilstedeværelse vil bli registrert. For at studenten skal kunne framstille seg til eksamen må arbeidskravet være godkjent senest 3 uker før eksamen.

## Evaluering av studiet

For å kunne tilby en aktuell og relevant utdanning av god kvalitet er vi avhengige av studentenes tilbakemeldinger. Evaluering gjennomføres i henhold til høyskolens kvalitetssystem.

## Tilbakemelding underveis

Det gis tilbakemelding via e-post mellom samlingene.

## Vurdering

### Mappevurdering

Mappen består av fire prosjekter. Det første prosjektet vil bli gitt etter første samling, det andre etter andre samling osv. Prosjektene skal gjennomføres individuelt eller i grupper på to og to. Det settes individuell karakter, også ved deltakelse i grupper på to.

Besvarelsen vurderes av en intern og en ekstern sensor. Det benyttes gradert karakterskala fra A - F, der A er beste og E er dårligste bestått karakter. F er ikke bestått.

### Plagiatkontroll

Alle skriftlige arbeidskrav og eksamensoppgaver kan plagiatkontrolleres. Plagiering og avskrift av faglitteratur og andre skriftlige arbeider uten korrekt bruk av referanser/kilder vil bli vurdert som forsøk på fusk. Se for øvrig [Forskrift om eksamen og studierett ved Høgskolen i Østfold](#)

## Litteratur

Studiet baserer seg i hovedsak på forelesningsmateriale og webressurser.

Anbefalt litteratur:

- Tom. M. Mitchell, Machine Learning, 1997, McGraw-Hill Higher Education; ISBN: 0070428077
- M. Bishop, Neural Networks for Pattern Recognition, 1996.
- H. Witten, E. Frank and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 2011.
- A. Ng, Machine Learning, Stanford University, online course.
- D. Koller, Probabilistic Graphical Models, Stanford University, online course.
- G. Hinton, Neural Networks for Machine Learning, University of Toronto, online course.
- M. Collins, Neural language Processing, Columbia University, online course.

## Jobb og videre studier

Personer med kompetanse innenfor analyse av store datamengder vil i en økende digitalisert verden bli svært ettertraktede ressurser.

Samfunnet er og blir, i større og større grad digitalisert; flere og flere av våre aktiviteter foregår på digitale arenaer. Fra kontakt med venner, familie og kolleger gjennom sosiale medier, til levering av selvangivelse, registrering av strømforbruk og bruk av medisinske målesystemer. Behovet for analyse av store datamengder blir bare større og større i dagens samfunn.

## Studieplanen er godkjent og revidert

### Studieplanen er godkjent

Dekan Beathe Due, 15.05.15

### Studieplanen gjelder for

Høsten 2015

## Studiemodell

Denne studiemodellen har en ny utforming. [Fortell oss hva du synes om den](#)

Høst 2015

Obl. emne



# ITVBIGD115 Analyse av Big Data (Høst 2015)

## Fakta om emnet

**Studiepoeng:** 15

**Ansvarlig avdeling:** Avdeling for informasjonsteknologi

**Undervisningsspråk:** Engelsk

**Varighet:** ½ år

## Innholdsfortegnelse

- Emnet er tilknyttet følgende studieprogram

## Emnet er tilknyttet følgende studieprogram

Obligatorisk emne i studiet Analyse av Big Data (15 studiepoeng). Studiet har kun ett emne, se studieplanen for nærmere beskrivelse.