

Kursplan – Data scientist, Utbildnings nr YH01458 - 2021 - 2		
Kurs: Maskininlärning	Poäng: 30 Yhp	Utgåva: Utgåva X
Framtagen av UL granskad av RUC:	Språk: Svenska / Engelska	Datum:
Förkunskaper: Inga förkunskaper	Granskad/Fastställd av: LG	S. 1/2

Kursens huvudsakliga innehåll

Kursen syftar till att ge den studerande specialiserade kunskaper i att förstå och använda maskininlärningstekniker för att kunna fatta datadrivna beslut. Under kursen kommer den studerande att lära sig att göra modellval, modellanpassning samt modellutvärdering i övervakade och oövervakade inlärningsproblem

Kursen omfattar följande moment:

- Statistik och maskininlärning
- Modellval, modellanpassning och modellutvärdering
- Övervakade och oövervakade inlärningsproblem
- Regression
- Klassificering
- Beslutsträd
- Ensemble learning
- Klustering
- Förståelse av datans betydelse

Kursens mål/läranderesultat

Målet med kursen är att den studerande ska kunna modellera data med hjälp av maskininlärningsmodeller. Detta innefattar b.l.a. att göra prediktioner, att förstå hur en anpassar en modell utifrån data och att kunna utvärdera modelleringsval och modeller.

Efter genomförd kurs ska den studerande kunna:

Kunskaper:

1. Redogöra för övervakad inlärningsmodellering såsom regression-och klassificeringsmodeller
2. Redogöra för oövervakad inlärnings modellering
3. Förklara hur en kan utvärdera modeller

Färdigheter:

4. Implementera metodik och modeller från maskininlärning med syfte att prediktera och förstå data
5. Diskutera motivera och argumentera för olika modelleringsval utifrån data

Kompetenser:

Inga kompetenser

Former för undervisning

Kursen kommer att genomföras med blended learning med inspelningar och aktiva lektioner. Under

kursens gång erbjuds även s.k. Open Office Hours, där studenterna har ytterligare möjlighet att få hjälp av kursledarna genom att ställa frågor.

Former för kunskapskontroll

Examination kommer att ske genom:

- 1 individuell inlämningsuppgift (IG/G/VG)
- 1 gruppuppgift vilken ska redovisas muntligt (IG/G)

Betygsskala

Följande betygsskala tillämpas:

VG = Väl Godkänd, G = Godkänd, IG = Icke Godkänd

Läranderesultat	Inlämningsuppgift (VG)	Gruppuppgift (G)
1	X	
2	X	
3	X	
4	X	X
5	X	X

Principer för betygssättning

För betyget Godkänd ska den studerande

- Ha nått samtliga läranderesultat för kursen

För betyget Väl Godkänd ska den studerande:

- Uppnått kraven för betyget Godkänd
- I en skriftlig rapport lösa ett problem genom att implementera metoder och modeller från maskininlärning på ett fördjupat sätt med hög säkerhet.
- Redogöra för och kritiskt diskutera modellval, modellenpassning och modellutvärdering

Icke Godkänd ges till studerande som har fullföljt kursen men inte nått alla mål för kursen.

Utbildare

Kursansvarig examinator: Antonio Prgomet.

E-post: Omniway.

Tillgänglighet: Möjligheten att ställa frågor och diskutera med utbildaren sker på lektionstid och vid behov via mejl funktionaliteten på Omniway. Försök att nyttja lektionstiden framför mejl för att kontakta mig.

Kursmaterial

Typ av material	Kommentar
Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems av Aurelien Geron (andra upplagan). 2:a upplagan används. ISBN: 9781492032649 Boken finns i pdf format.	Kursboken har en GitHub sida med kod: https://github.com/ageron/handson-ml2
Kursens GitHub sida: https://github.com/AntonioPrgomet/machine_learning_ds23	Allt kursmaterial finns på GitHub länken.
Inspelade kursvideos: https://www.youtube.com/playlist?list=PLgzaMbMPEHEX9Als3F3sKKXexWnyEKH45	
Via skolan har du gratis tillgång till DataCamp: https://www.datacamp.com/groups/shared_links/1041b274b6fce57c4792a0eb9aff16edbc744843654b248f7cc2779894f6697b Gå in på en kurs och kolla på t.ex. kapitel 3, om det funkar så funkar din DataCamp (kapitel 1 i kurserna är gratis så därför behöver man kolla t.ex. kapitel 3 i en kurs).	DataCamp online kurser. Registrera dig i DataCamp med din skolmejl (t.ex. malte.maltesson@utb.ec utbildning.se) annars fungerar det inte.
8 videor om studieteknik, de kommer hjälpa dig mycket under dina studier och även senare i arbetslivet när du skall lära dig nya saker. https://www.youtube.com/watch?v=gSbpRjxYq24&list=PLA09CC1B5671827AD	

Kunskapskontroll - Regler

Resultat och betyg registreras senast inom 10 arbetsdagar från deadline.

Om man inte kan utföra examinationen på utsatt deadline (till exempel på grund av allvarlig sjukdom) eller behöver utföra en komplettering så är "andra" examinationstillfället senast en vecka efter att den studerande fått sin uppgift rättad eller utsatt deadline om man inte gjort examinationen. Du kan lämna in examinationen när du vill under denna period.

Det "tredje" (och sista) examinationstillfället är senast tre veckor efter att den studerande fått sin uppgift rättad vid senaste tillfället eller senaste deadline om ingen inlämning har gjorts. Har du missat samtliga examinationer måste du kontakta din utbildningsledare snarast.

Kunskapskontroll – Information

Kursen har två kunskapskontroller. Se veckoplaneringen för när de lämnas in.

Upplägg på Föreläsningar / Lektioner

Lektionerna kommer fokusera på genomgångar och lösningar av uppgifter. Uppstår det frågor under självstudierna så skriv ned dem så tar vi upp dem på lektionstid. Viktigt att man *inför* lektionerna arbetat med det material som förväntas enligt planeringen.

Schema:

	Förmiddag: 08.15 – 12.00.	Eftermiddag: 13.15 – 17.00
Måndag	Egenstudier för samtliga orter	Egenstudier för samtliga orter
Tisdag	Helsingborg / Malmö	Stockholm
Onsdag	Göteborg	Helsingborg / Malmö
Torsdag	Stockholm	Göteborg
Fredag	Egenstudier för samtliga orter	Egenstudier för samtliga orter

Veckoplanering – Vad skall jag göra varje arbetsdag?

I detta avsnitt så framgår i detalj vad som skall göras varje dag. Viktigt att du följer schemat.

	Kursvecka 1 (v.7) - Introduktion och skapa sig en överblick (Mycket gås igenom)
Mån	<p>Kolla video "Introduktion till Maskininlärning": https://www.youtube.com/watch?v=zORv5vxrwok&list=PLgzaMbMPEHEEx9Als3F3sKKXexWnyEKH45&index=2&t=1719s</p> <p>Läs kapitel 1 i kursboken. - Vill du arbeta med uppgifter så prioritera dessa: 1, 2, 3, 4, 7, 12, 15, 16, 17, 19.</p>
Tis	<p>Lektion kl: 08.15 - 12.00.</p> <p>EM: Kolla följande videos "Klassificering": https://www.youtube.com/watch?v=-QzFLifZgNw&list=PLgzaMbMPEHEEx9Als3F3sKKXexWnyEKH45&index=2</p> <p>"Linjär Regression & Logistisk Regression": https://www.youtube.com/watch?v=0f1tWgarTkQ&list=PLgzaMbMPEHEEx9Als3F3sKKXexWnyEKH45&index=4&t=1912s</p> <p>"Regression - Kod Exempel": https://www.youtube.com/watch?v=T24GpzKHS14&list=PLgzaMbMPEHEEx9Als3F3sKKXexWnyEKH45&index=4&t=101s&pp=gAOBiAOB</p> <p>"Klassificering - Kod Exempel": https://www.youtube.com/watch?v=MNELnQx7L8M&list=PLgzaMbMPEHEEx9Als3F3sKKXexWnyEKH45&index=6&t=2s</p> <p>Läs kapitel 2 i kursboken. - Vill du göra några uppgifter så prioritera dessa: 1, 4.</p>

Ons	<p>FM:</p> <p>Kolla följande videos</p> <p>"Klassificering": https://www.youtube.com/watch?v=-QzFLifZgNw&list=PLgzaMbMPEHEx9Als3F3sKKXexWnyEKH45&index=2</p> <p>"Linjär Regression & Logistisk Regression": https://www.youtube.com/watch?v=0f1tWgarTkQ&list=PLgzaMbMPEHEx9Als3F3sKKXexWnyEKH45&index=4&t=1912s</p> <p>"Regression - Kod Exempel": https://www.youtube.com/watch?v=T24GpzKHSI4&list=PLgzaMbMPEHEx9Als3F3sKKXexWnyEKH45&index=4&t=101s&pp=gAQBiAQB</p> <p>"Klassificering - Kod Exempel": https://www.youtube.com/watch?v=MNELnQx7L8M&list=PLgzaMbMPEHEx9Als3F3sKKXexWnyEKH45&index=6&t=2s</p> <p>Läs kapitel 2 i kursboken.</p> <p>- Vill du göra några uppgifter så prioritera dessa: 1, 4.</p> <p>Lektion kl: 13.15 - 17.00.</p>
Tor	<p>Läs kapitel 3 i kursboken.</p> <p>- Om du har tid över jobba med uppgift 3 "Titanic Dataset".</p>
Fre	<p>Läs kapitel 4 i kursboken.</p>

	Kursvecka 2 (v.8) - Fokus på programmering
Mån	Gör följande DataCamp kurs kopplat till supervised learning (regression och klassificering): https://app.datacamp.com/learn/courses/supervised-learning-with-scikit-learn - Oroa dig inte om du inte förstår alla detaljerna. Det tar tid att smälta och vi kommer gå igenom flera koncept i princip under hela kursen.
Tis	Lektion kl: 08.15 - 12.00. - Utdelning av kunskapskontroll 1. EM: Arbeta med kunskapskontroll 1.
Ons	FM: Arbeta med kunskapskontroll 1. Lektion kl: 13.15 - 17.00. - På lektionen kommer vi diskutera och presentera det som gjorts på kunskapskontroll 1. Lämna in kunskapskontroll 1 på Omniway efter lektionen. Deadline kl: 23.59.
Tor	Kolla på videon "Design Principles of Scikit-Learn: Theory and Code Demonstrations": https://youtu.be/Fy9o4pXRVzs och jobba igenom de olika skripten tillgängliga på GitHub, länk till GitHub finns i videobeskrivningen.
Fre	Gör följande DataCamp kurs (endast del 1 – 3, del 4 om Support Vector Machines görs senare): https://app.datacamp.com/learn/courses/linear-classifiers-in-python

	Kursvecka 3 (v.9) - Ännu mer programering
Mån	<p>Kolla video "Support Vector Machines": https://www.youtube.com/watch?v=NzGiMYinmog&list=PLgzaMbMPEHEx9Als3F3sKKXexWnyEKH45&index=7</p> <p>Läs kapitel 5 i kursboken. - Avsnitten under "Under the Hood" kan du läsa översiktligt för att se helhetsbilden, men detaljerna är överkurs och ingår inte i kursen. - Jobba med följande uppgifter: 1, 2, 3, 6 från kapitel 5.</p>
Tis	<p>Lektion kl: 08.15 - 12.00.</p> <p>EM: Gör följande DataCamp kurs (OBS: endast del 4 om Support Vector Machines): https://app.datacamp.com/learn/courses/linear-classifiers-in-python</p> <p>Jobba med uppgift 9 från kapitel 5 (Kunskapskontroll 2 är baserad på den uppgiften så bra träning).</p>
Ons	<p>FM: Gör följande DataCamp kurs (OBS: endast del 4 om Support Vector Machines): https://app.datacamp.com/learn/courses/linear-classifiers-in-python</p> <p>Jobba med uppgift 9 från kapitel 5 (Kunskapskontroll 2 är baserad på den uppgiften så bra träning).</p> <p>Lektion kl: 13.15 - 17.00.</p>
Tor	Repetition.
Fre	<p>Kolla video om "Beslutsträd": https://www.youtube.com/watch?v=X-uXL4x2Ap0&list=PLgzaMbMPEHEx9Als3F3sKKXexWnyEKH45&index=8</p> <p>Läs kapitel 6 i kursboken. - Vill du göra några uppgifter så prioritera uppgifterna: 3, 4, 7.</p>

	Kursvecka 4 (v.10) - Teori & Programmering
Mån	<p>Kolla video "Ensemble Learning & Random Forest": https://www.youtube.com/watch?v=-kErwtsQRKk&list=PLgzaMbMPEHEX9Als3F3sKKXexWnyEKH45&index=9</p> <p>Läs kapitel 7 i kursboken (allt under avsnittet "Boosting" kan du skippa då det inte ingår i kursen, alternativt så läser du det kort och översiktligt för att ha "hört det" någon gång).</p> <p>- Vill du göra några uppgifter så prioritera uppgifterna: 1, 2.</p>
Tis	<p>Lektion kl: 08.15 - 12.00.</p> <p>EM: Repetition.</p>
Ons	<p>FM: Repetition.</p> <p>Lektion kl: 13.15 - 17.00.</p>
Tor	<p>Kolla video "Dimensionsreducering - PCA": https://www.youtube.com/watch?v=VHZRRXEZijA&list=PLgzaMbMPEHEX9Als3F3sKKXexWnyEKH45&index=10</p> <p>Läs kapitel 8 i kursboken.</p> <p>- Jobba med följande uppgifter: 1, 2, 4.</p>
Fre	<p>Kolla video "Unsupervised Learning Techniques" (OBS: knappt 3 h lång): https://www.youtube.com/watch?v=PjeAU1hrGk&list=PLgzaMbMPEHEX9Als3F3sKKXexWnyEKH45&index=11</p> <p>Läs kapitel 9 i kursboken.</p> <p>- Allt efter "Gaussian Mixtures" ingår inte i kursen, skippa det.</p> <p>- Jobba med följande uppgifter: 1, 2.</p>

	Kursvecka 5 (v.11) - Utdelning av kunskapskontroll 2
Mån	Repetition.
Tis	Lektion kl: 08.15 - 12.00. - Utdelning av kunskapskontroll 2.
Ons	FM: Lektion kl: 13.15 - 17.00.
Tor	Kolla på videon "Introduction to Streamlit": https://www.youtube.com/watch?v=ggDa-RzPP7A&list=PLgzaMbMPEHEX9Als3F3sKKXexWnyEKH45&index=12
Fre	Arbeta med kunskapskontroll 2.

	Kursvecka 6 (v.12) - Slutföra kunskapskontroll 2
Mån	Arbeta med kunskapskontroll 2.
Tis	Lektion kl: 08.15 - 12.00. EM: Arbeta med kunskapskontroll 2.
Ons	FM: Arbeta med kunskapskontroll 2. Lektion kl: 13.15 - 17.00.
Tor	Arbeta med kunskapskontroll 2.
Fre	Deadline kunskapskontroll 2 kl: 23.59. Lämnas in på Omniway.