Web Intelligence

2DV515

Johan Hagelbäck



@johan.hagelback

Mats Loock



@mats

Vad är Web Intelligence?

- Idag har vi tillgång till extrema mängder data:
  - Rekommendationer för film, musik, böcker, …
  - Wikipedia
  - Produktrekommendationer och köp
  - Väderdata
  - …
- Web Intelligence kan kortfattat summeras så här:
  - Hur kan vi lära oss och dra slutsatser av data för att förbättra befintliga produkter/tjänster eller skapa helt nya produkter/tjänster?

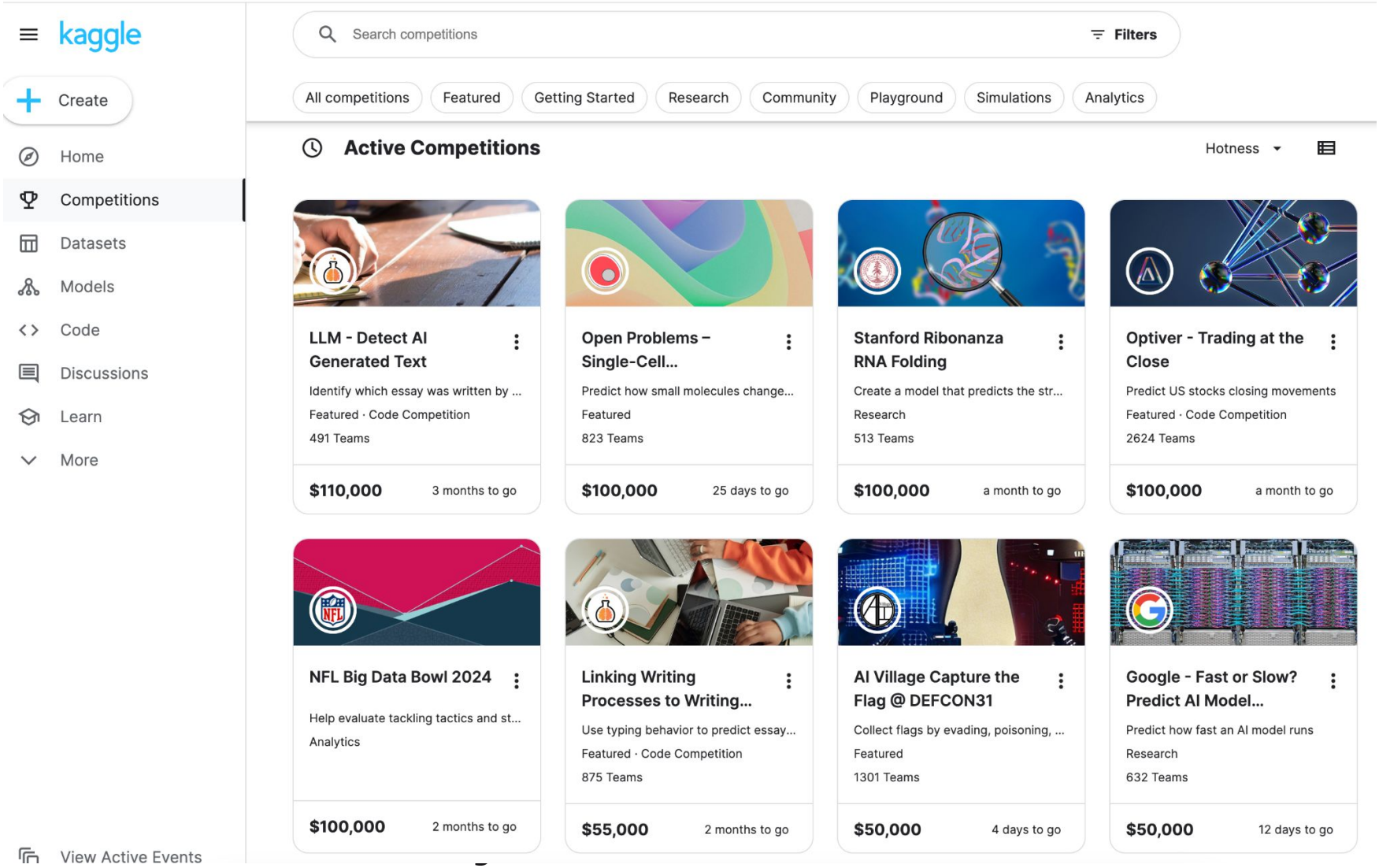


Netflix

- Netflix rekommenderar filmer eller TV serier som du inte redan sett och som kanske kan passa dig
- År 2006 gick Netflix ut med en tävling där målet var att förbättra träffsäkerheten för deras rekommendationer med 10%
- I prispotten låg $1 miljon för de som först lyckades nå detta
- Ett år senare var bästa laget uppe i 7% förbättring
- Det dröjde till september 2009 innan ett lag, *BellKor's Pragmatic Chaos team*, uppnådde 10%
- Det finns uppenbarligen stora fördelar med bättre rekommendationssystem



Kaggle


kaggle
Search competitions
Filters
Create
All competitions
Featured
Getting Started
Research
Community
Playground
Simulations
Analytics
Home
Competitions
Datasets
Models
Code
Discussions
Learn
More
Active Competitions
Hotness
LLM - Detect AI Generated Text
Identify which essay was written by ...
Featured · Code Competition
491 Teams
$110,000
3 months to go
Open Problems – Single-Cell...
Predict how small molecules change...
Featured
823 Teams
$100,000
25 days to go
Stanford Ribonanza RNA Folding
Create a model that predicts the str...
Research
513 Teams
$100,000
a month to go
Optiver - Trading at the Close
Predict US stocks closing movements
Featured · Code Competition
2624 Teams
$100,000
a month to go
NFL Big Data Bowl 2024
Help evaluate tackling tactics and st...
Analytics
$100,000
2 months to go
Linking Writing Processes to Writing...
Use typing behavior to predict essay...
Featured · Code Competition
875 Teams
$55,000
2 months to go
AI Village Capture the Flag @ DEFCON31
Collect flags by evading, poisoning, ...
Featured
1301 Teams
$50,000
4 days to go
Google - Fast or Slow? Predict AI Model...
Predict how fast an AI model runs
Research
632 Teams
$50,000
12 days to go
View Active Events

Google

- Innan Google startade utgick sökmotorer endast från innehållet på webbsidor
- Larry Page och Sergei Brin tog en helt annan väg
- De rankade webbsidor inte bara utifrån innehållet, utan även utifrån hur populär en webbsida är
- Detta ledde till en betydligt bättre sökmotor som presenterade mer relevanta resultat
- … och Google startades



Web Intelligence

- Detta är några exempel på vad Web Intelligence är och hur viktigt det är
- Web intelligence täcker in många olika områden:
  - AI
  - Maskininlärning
  - Visualisering
  - Informationsteknologi
  - Statistik
  - NLP (Natural Language Processing)
  - …



Om kursen



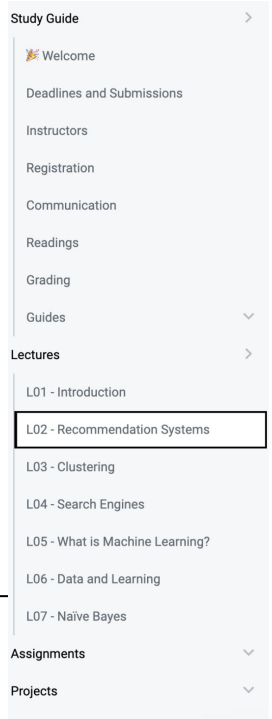
Föreläsningar

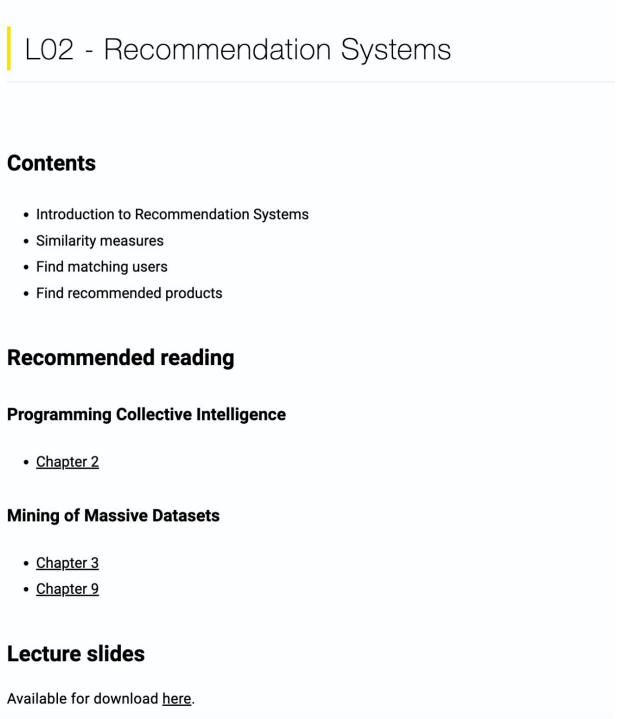
- Kursen består av 7 förinspelade föreläsningar:
  1. Introduktion
  2. Rekommendationssystem
  3. Klustring
  4. Sökmotorer
  5. Vad är maskininlärning?
  6. Data och inlärning
  7. Naïve Bayes algoritmen
- Föreläsningarna har några år på nacken, men de behandlar grundläggande tekniker som fortfarande är relevanta och viktiga



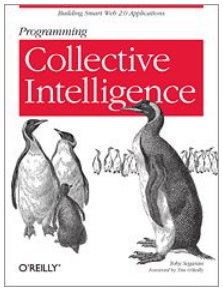
Föreläsningar

- Titta på föreläsningarna i din egen takt
- Läs vidare för att få en djupare förståelse för de olika delarna

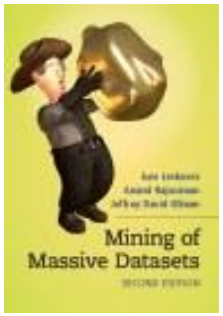




Literatur



- Programming Collective Intelligence
- Seagaran, Toby (2008)
- Har rätt många år på nacken, men många algoritmer som vi använder än idag förklaras väldigt bra



- Mining of Massive Datasets
- Leskovec, Jure & Rajaraman, Anand & Ullman, Jeffrey David (2014)
- Gratis e-bok. Förklarar teoretiska koncept på ett bra sätt med många exempel

- … och en del online-resurser

Inlämningar

- Fyra inlämningsuppgifter och ett projekt
- Muntliga examinationer online via Zoom
  https://coursepress.lnu.se/2dv515/zoom
- Examinationstider bokas på MyMoodle
- Inlämning via Gitlab
- Tydliga mål för olika betyg - välj vad du vill sikta på!



Inlämningsuppgifter

- **A1**
  - Rekommendationssystem för filmer
  - Klient/server webapplikation med REST-ish interface
- **A2**
  - Klustring av bloggar
  - Klient/server webapplikation med REST-ish interface
- **A3**
  - Sökmotor för ett urval Wikipedia-sidor
  - Klient/server webapplikation med REST-ish interface
- **A4**
  - Maskinlärning med Naïve Bayes-algoritmen

Du kan välja vilket programspråk och klient/server ramverk du vill!



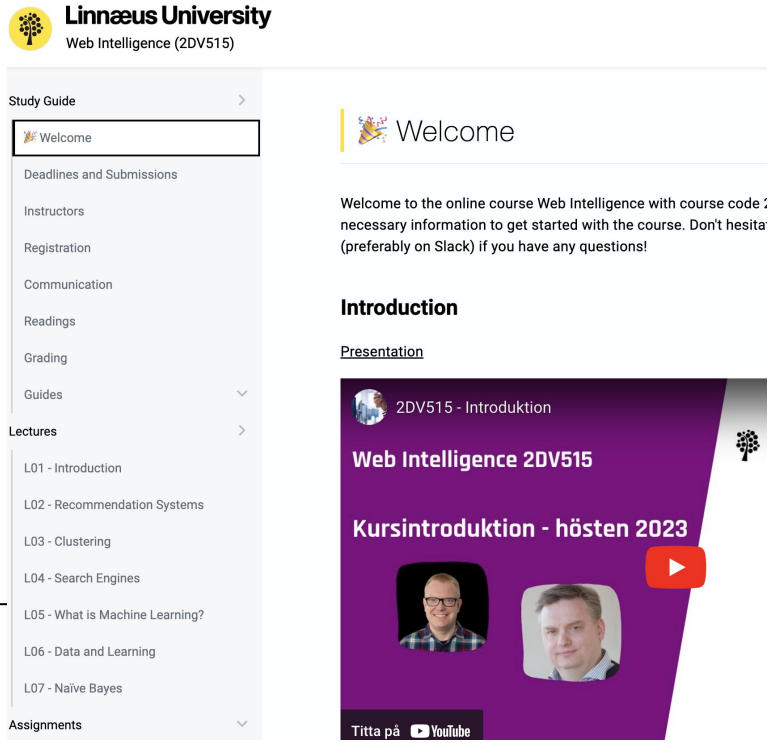
Projekt

- Det finns fyra olika projekt att välja mellan:
  - Rekommendationssystem
  - Klustring
  - Web scraping och sökmotor
  - Maskininlärning
- Projekten bygger på inlämningsuppgifterna
- Välj det projekt du tycker verkar mest intressant!



Kurssidor

- MyMoodle används endast för att boka tider för examination
- Alla föreläsningar, instruktioner etc. finns på Coursepress



Betygssättning

- En totalpoäng för kursen beräknas enligt följande:
  $Score_F = (Score_{A1} + Score_{A2} + Score_{A3} + Score_{A4} + 3.5 * Score_{Project}) / 7.5$
- Poängen avrundas till närmaste heltal och tabellen nedan visar vilket betyg du får:

Grade	A	B	C	D	E
Score	5	4	3	2	1

Betygssättning

- Exempel:
  - Du har fått följande betyg:

	A1	A2	A3	A4	Projekt
Betyg	C	A	B	A	B

  - Din slutpoäng blir då:
    (3 + 5 + 4 + 5 + 3.5 * 4) / 7.5 = 4.13 ≈ 4
  - … vilket enligt tabellen blir **B**

Kommunikation

- Frågor, diskussioner etc. tas med fördel på Slack
- Generella frågor om kursen eller ämnet #2DV515
- Specifika frågor om dina inlämningar eller betyg tas via Slack DM
- Håll kolla på deadlines och tveka inte att ställa frågor!

