



What is the case about?



Picking a movie to watch is hard. Let's build

a movie recommender system

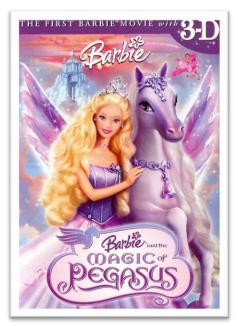


What is the goal?

- The goal is to be able to recommend **similar movies** for a given movie
- For instance: If you like *My Little Pony*, you might also like *Barbie and the Magic of Pegasus*









What data will you use?

You will receive a dataset with a few thousand movies exported from **IMDB** with

various columns:

- Title
- Description
- Language
- Popularity
- Runtime
- Genres
- Release date
- Revenue
- **Etc**

title	▼ description ▼
	Led by Woody, Andy's toys live happily in his room until Andy's birthday brings
	Buzz Lightyear onto the scene. Afraid of losing his place in Andy's heart, Woody
	plots against Buzz. But when circumstances separate Buzz and Woody from their
Toy Story	owner, the duo eventually learns to put aside their differences.
	When siblings Judy and Peter discover an enchanted board game that opens the
	door to a magical world, they unwittingly invite Alan an adult who's been
	trapped inside the game for 26 years into their living room. Alan's only hope
	for freedom is to finish the game, which proves risky as all three find themselves
Jumanji	running from giant rhinoceroses, evil monkeys and other terrifying creatures.
	A family wedding reignites the ancient feud between next-door neighbors and
	fishing buddies John and Max. Meanwhile, a sultry Italian divorcée opens a
	restaurant at the local bait shop, alarming the locals who worry she'll scare the
	fish away. But she's less interested in seafood than she is in cooking up a hot
Grumpier Old Men	time with Max.
	Cheated on, mistreated and stepped on, the women are holding their breath,
	waiting for the elusive "good man" to break a string of less-than-stellar lovers.
	Friends and confidants Vannah, Bernie, Glo and Robin talk it all out, determined
Waiting to Exhale	to find a better way to breathe.
	Just when George Banks has recovered from his daughter's wedding, he receives
	the news that she's pregnant and that George's wife, Nina, is expecting too.
5.1. 6.1. 6.1. 6	He was planning on selling their home, but that's a plan that like George will
Father of the Bride Part II	have to change with the arrival of both a grandchild and a kid of his own.



What method will you use?

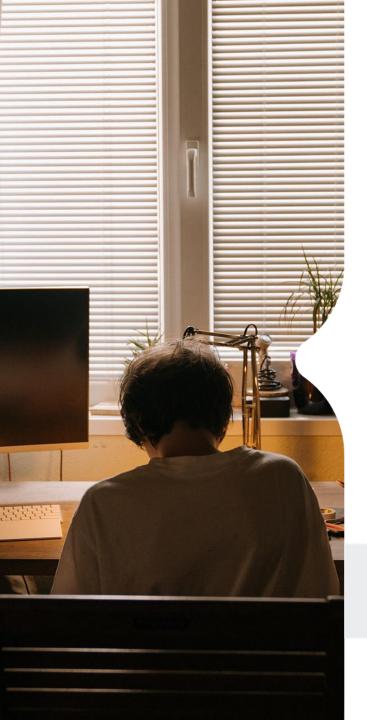
- There are many different methods for making recommendations. In this case we'll focus on using NLP methods based on movie descriptions
- You will split into teams of about 4 people. You will have access to a **Jupyter notebook** that performs these NLP methods and implements a basic version of a recommender
- Your job is to:
 - Select a movie
 - Generate recommendations
 - And improve upon this basic implementation
- After working on this, your team will give a short presentation of <u>3 minutes</u> to a jury



What should be in your presentation?

- In your presentation (3 min) you should:
 - **■** Present your current recommendation results
 - Share your **chosen movie** and its **recommended movies**
 - Discuss what **changes/tweaks** you made
 - **Convince the client of ideas for the future**
 - Pitch **suggestions** that could further improve the recommender
 - Suggest other **recommender methodologies** that could be used





Data preparation and modelling: an example pipeline

Goeie dag, me beeldscherm van Iphone is vervangen maar zelfde roze streep en nuu ??

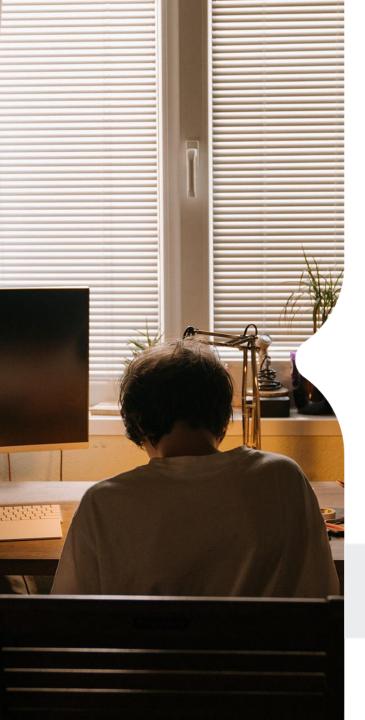












Data preparation and modelling: an example pipeline

Goeie dag, me beeldscherm van Iphone is vervangen maar zelfde roze streep en nuu??



goeie dag, me beeldscherm van iphone is vervangen maar zelfde roze streep en

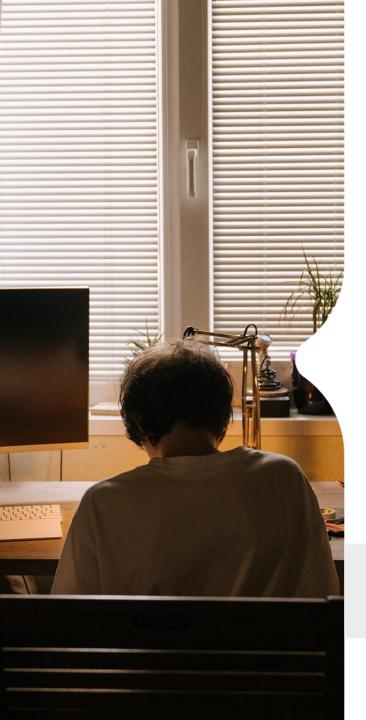












Data preparation and modelling: an example pipeline

Goeie dag, me beeldscherm van Iphone is vervangen maar zelfde roze streep en nuu??















The Distributional **Hypothesis**

"You shall know a word by the company it keeps." – Firth (1957)

Zit je **jack** aan de onderkant of bovenkant van je telefoon?

Mijn oortjes passen niet in de **jack** van mijn telefoon

Mijn audio **jack** is kapot, ik hoor geen geluid meer

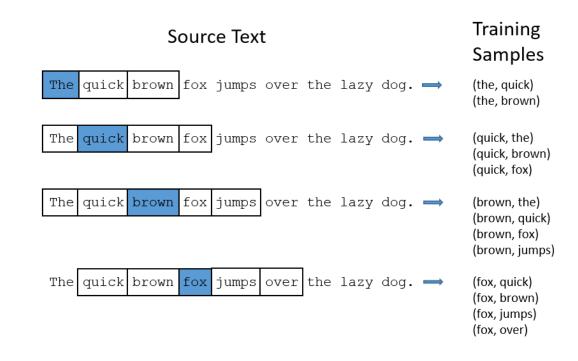


The meaning of the word is captured by the context



Word2Vec (Mikolov et al., 2013a)

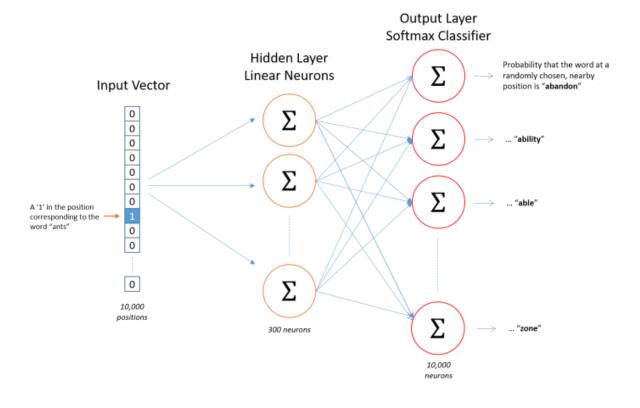
Create contextual embeddings by predicting surrounding words

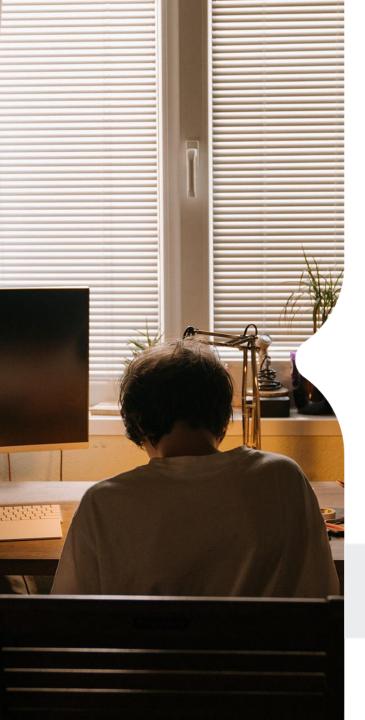




Word2Vec (Mikolov et al., 2013a)

f Shallow fully connected **neural network**





Data preparation and modelling: an example pipeline

goeie dag, me beeldscherm van iphone is vervangen maar zelfde roze streep en

}

goeie dag, ik heb roze streep op mijn scherm



1 2 7

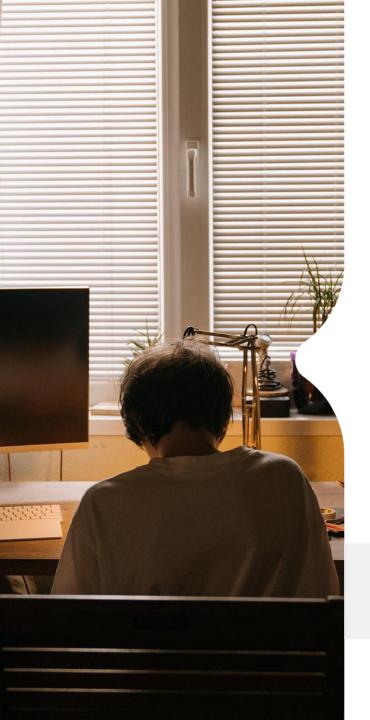












Data preparation and modelling: an example pipeline



+

Similarity: 0.78



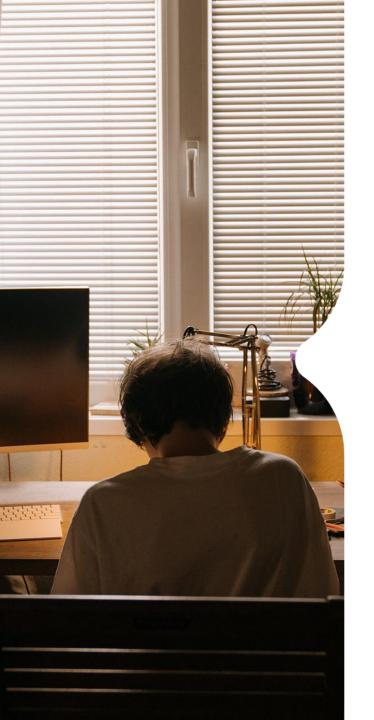












Data & Jupyter notebook can be found on Github



https://github.com/enniasuijkerbuijk/bb-nlp-case