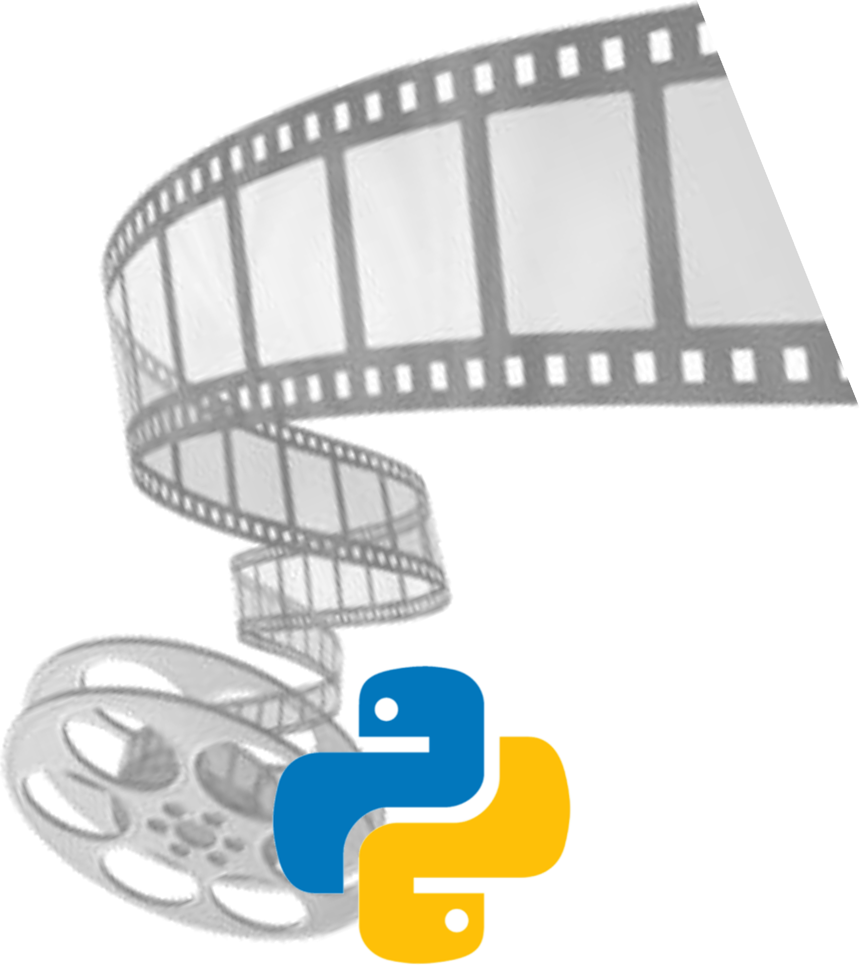
greatlearning
Learning for Life



Movie Recommendation System

Sampriti Chatterjee (Great Learning)

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Agenda

1 Why do we need data science?

2 What is Data science?

3 Life cycle of Data science

4 Why Python is so popular?

5 Install python

6 Statistical visualization on Python user

7 What is recommendation system?

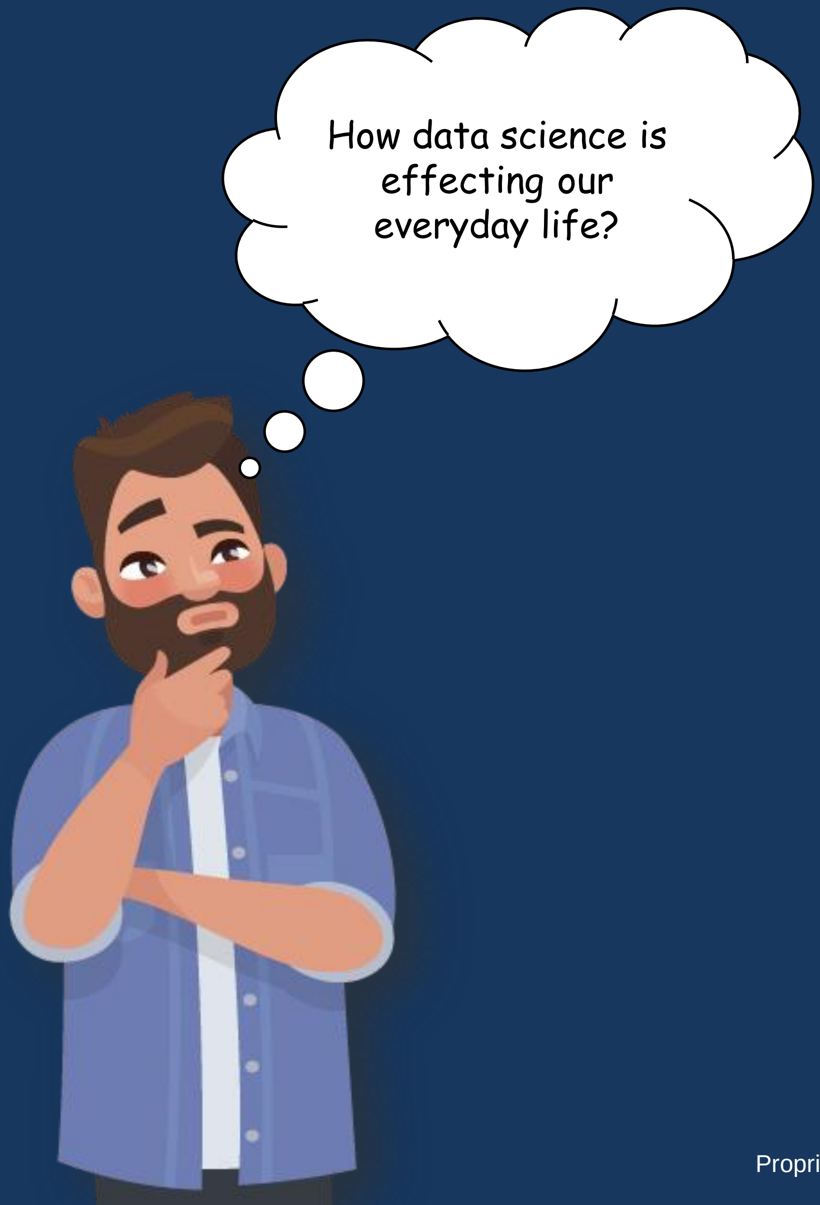
8 Types of recommendation system

9 Use case for UBCF

10 Use case for CBF

11 Demo: Movie recommendation system

Why do we need Data Science?



- In the past, we used to have data in a structured format but now as the volume of the data is increasing, so the number of structured data becomes very less, so to handle the massive amount of data we need data science techniques
- Those data can be used to get the proper business insights and the hidden trends from them.
- These insights helps the organization to predict the Future
- Using data science decision making can be faster and effective
- Helps to reduce the production cost
- Build model based on the data to give the ability to the machine to predicts on its own

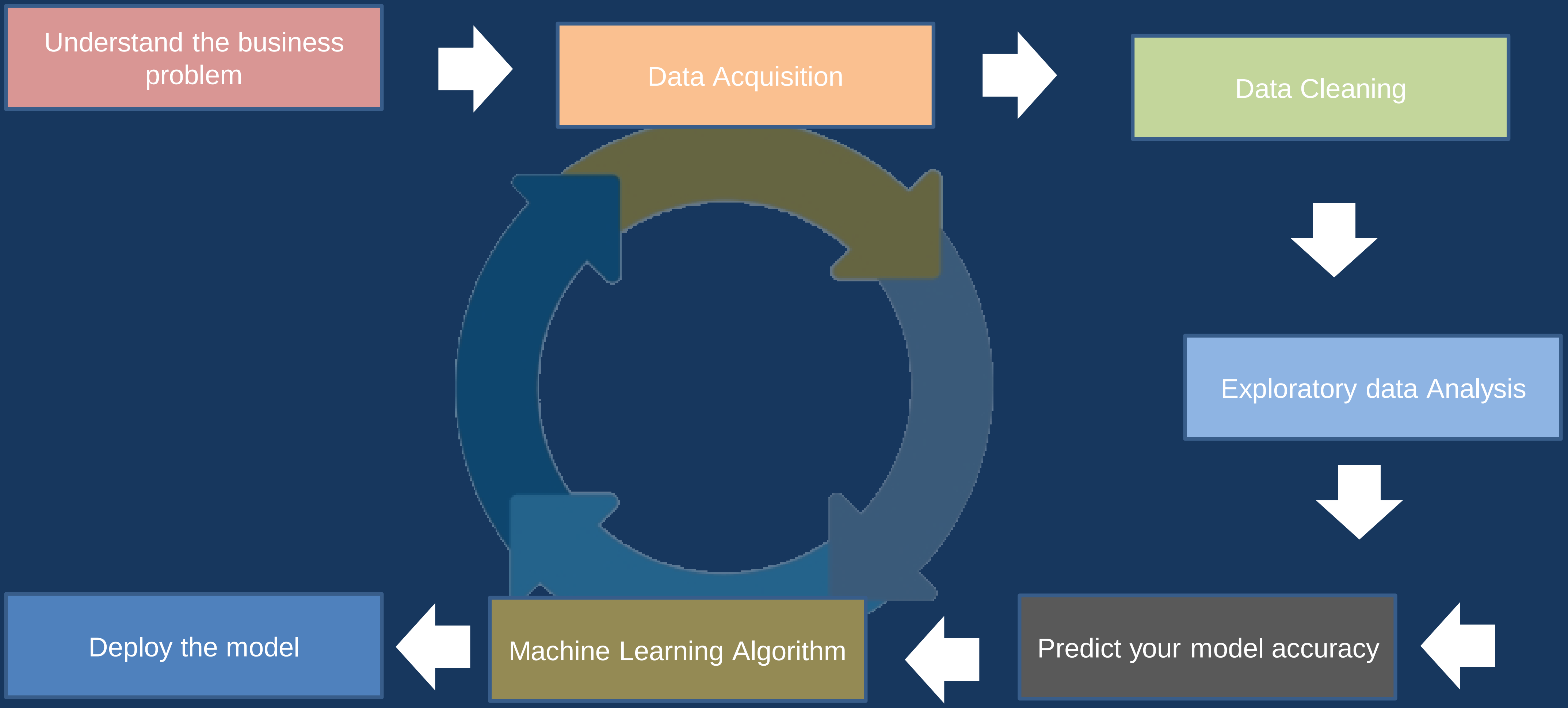
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

What is Data Science?

Data science is a process to get some meaningful information from the massive amount of data. In simple terms, read and study the data to get proper intuitive insights. Data Science is a mixture of various tools, algorithms, and machine learning and deep learning concepts to discover hidden patterns from the raw and unstructured data

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Life cycle of Data Science?



Understand the business problem

Data Acquisition

Data Cleaning

Exploratory data Analysis

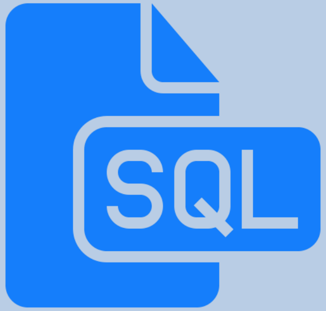
Predict your model accuracy

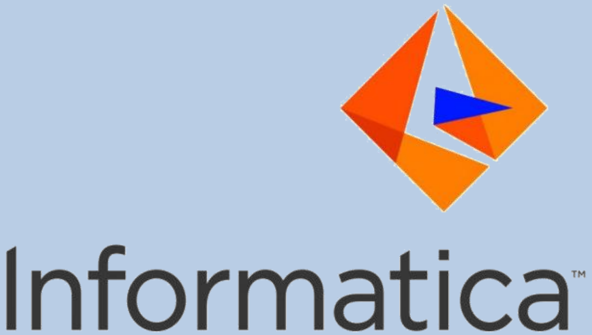
Machine Learning Algorithm

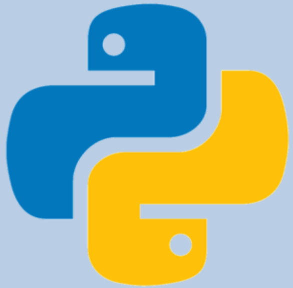
Deploy the model

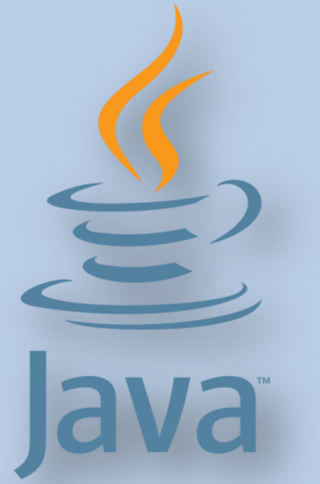
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

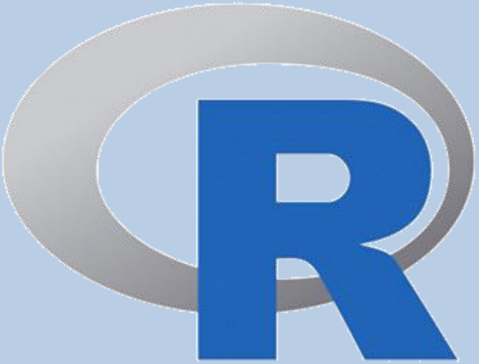
Most Popular Programming Languages For Data Science?

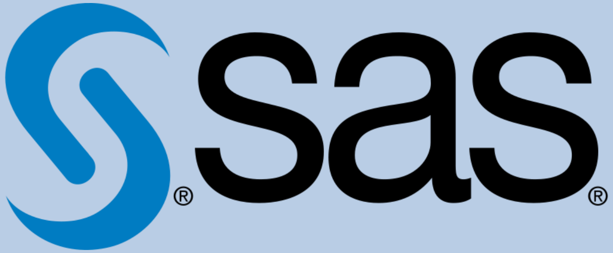


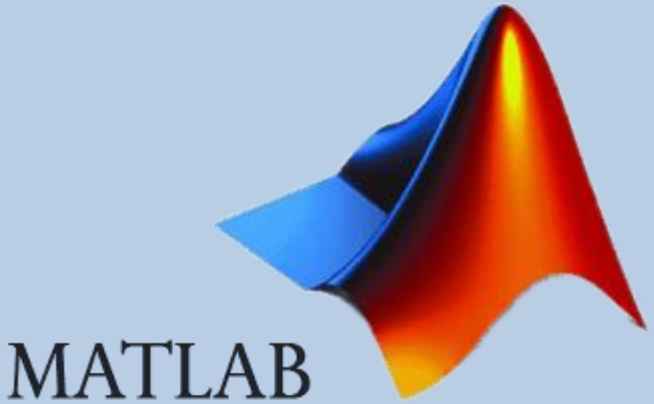








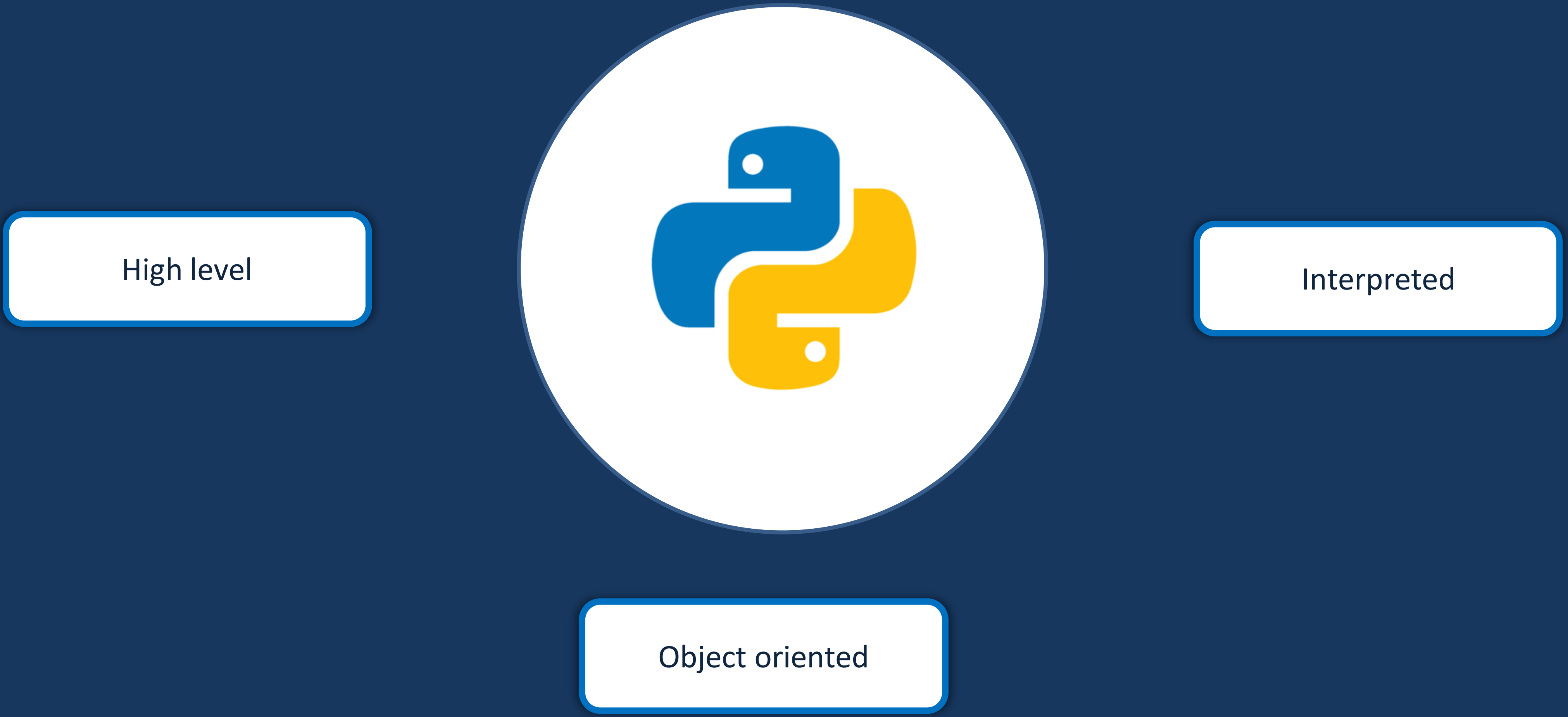




Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Introduction to Python

Python is a popular high level, object oriented and interpreted language



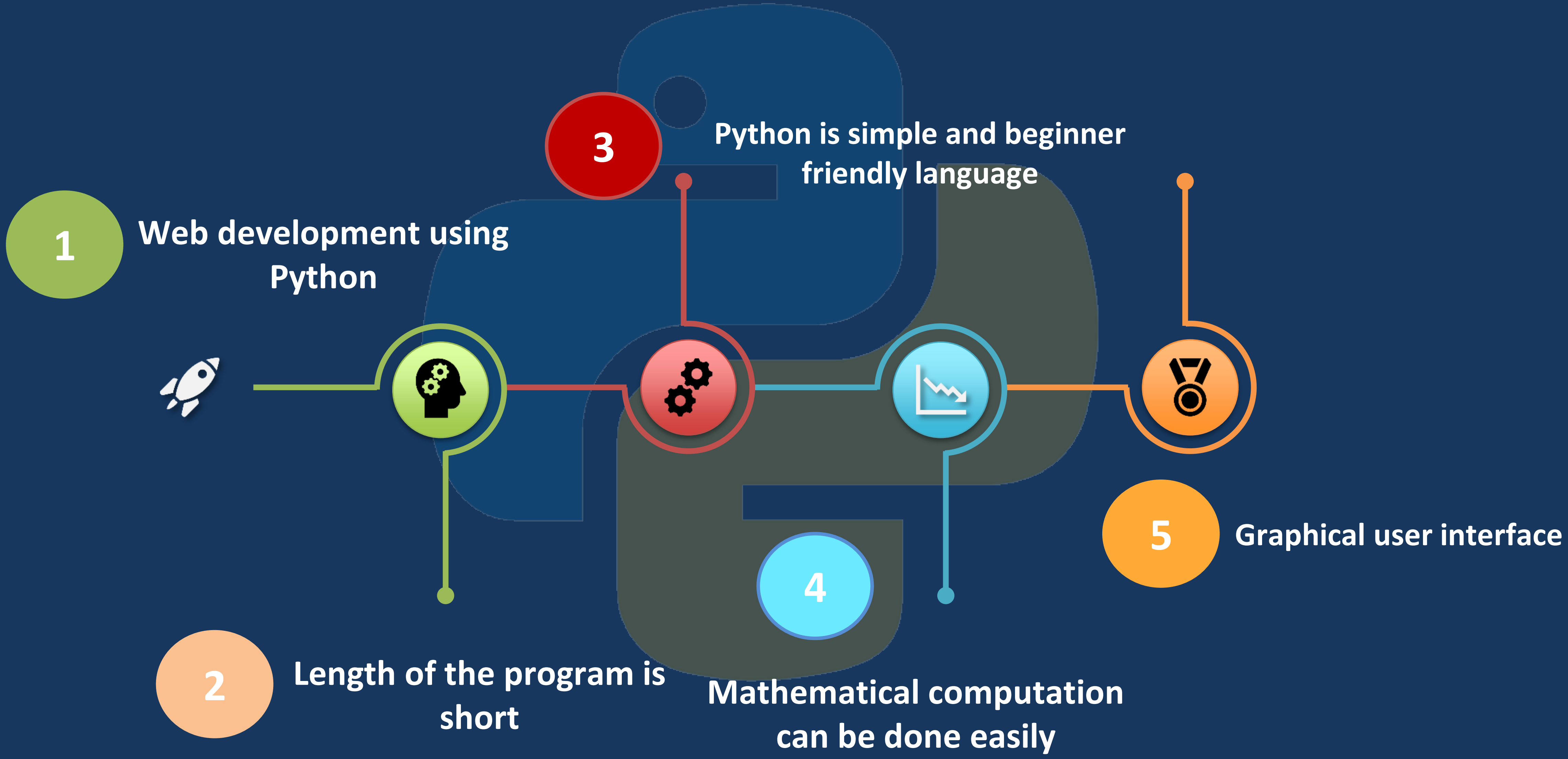
High level

Interpreted

Object oriented

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Why should you learn Python?



1 Web development using Python

2 Length of the program is short

3 Python is simple and beginner friendly language

4 Mathematical computation can be done easily

5 Graphical user interface

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Why Python is so popular?

1 Largest community for Learners and Collaborators

2 **Open source**

3 **Easy to learn and usable flexibility**

4 **Huge numbers of Python libraries and Frame work**

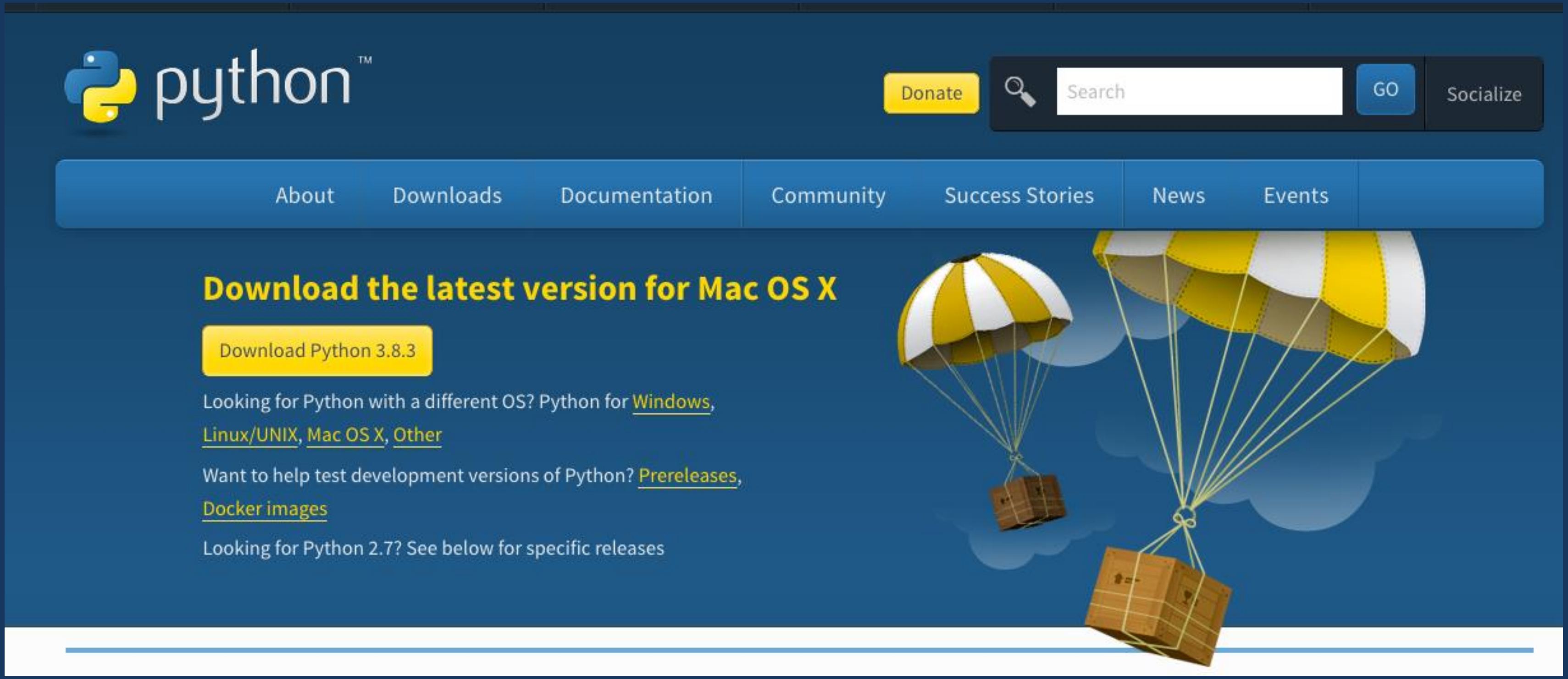
5 **Supports Big Data, Machine Learning and Cloud computing**

6 **Supports Automation**

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Installing Python

This is the site to install Python -> https://www.python.org/downloads/



Download the latest version for Mac OS X

Looking for Python with a different OS? Python for Windows, Linux/UNIX, Mac OS X, Other

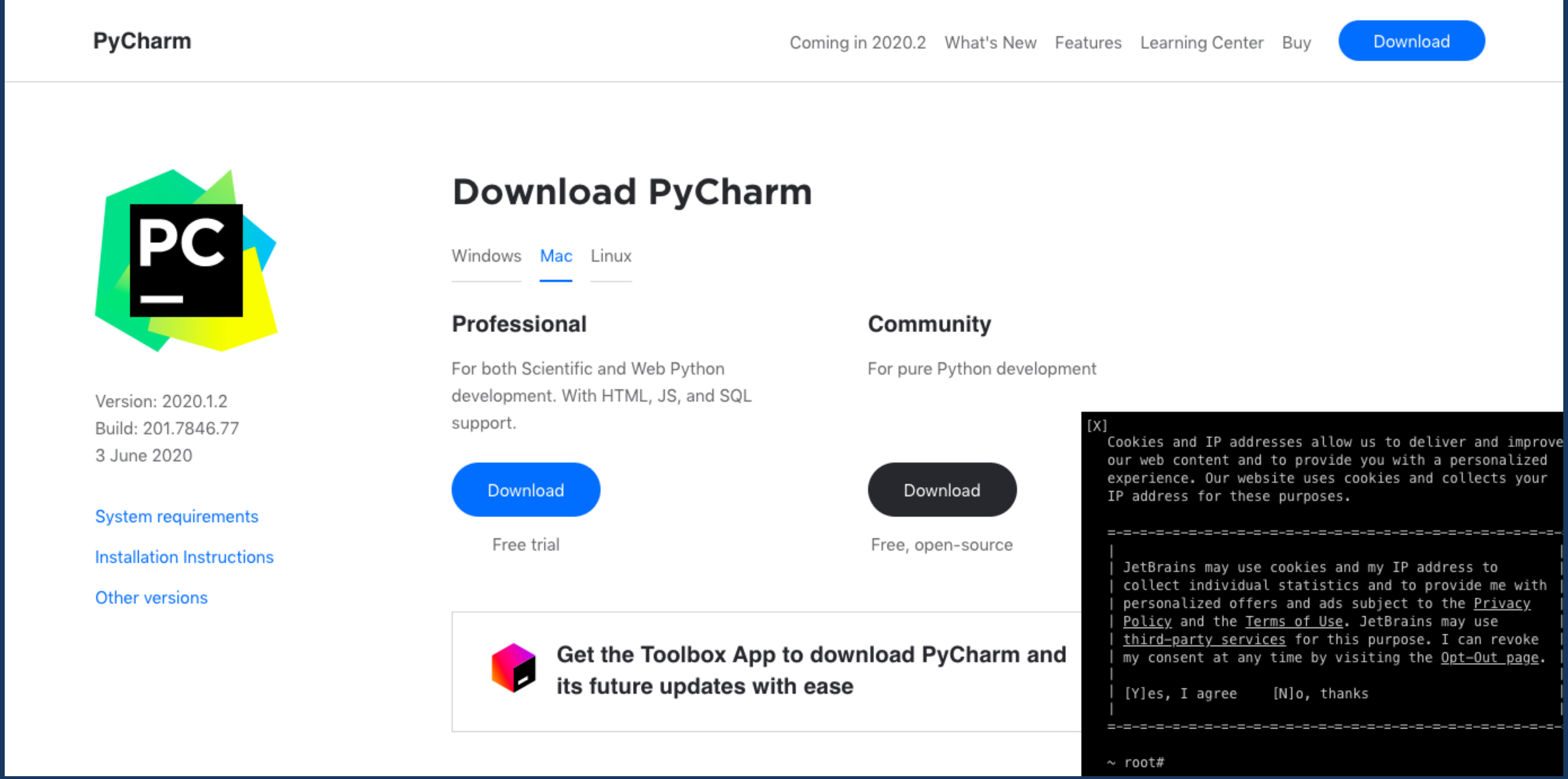
Want to help test development versions of Python? Prereleases, Docker images

Looking for Python 2.7? See below for specific releases

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Popular IDE for Python: Pycharm

Site to install Python ->
https://www.jetbrains.com/pycharm/download/#section=mac



Download PyCharm

Windows Mac Linux

Professional

For both Scientific and Web Python development. With HTML, JS, and SQL support.

Download

Free trial

Community

For pure Python development

Download

Free, open-source

Version: 2020.1.2
Build: 201.7846.77
3 June 2020

System requirements

Installation Instructions

Other versions

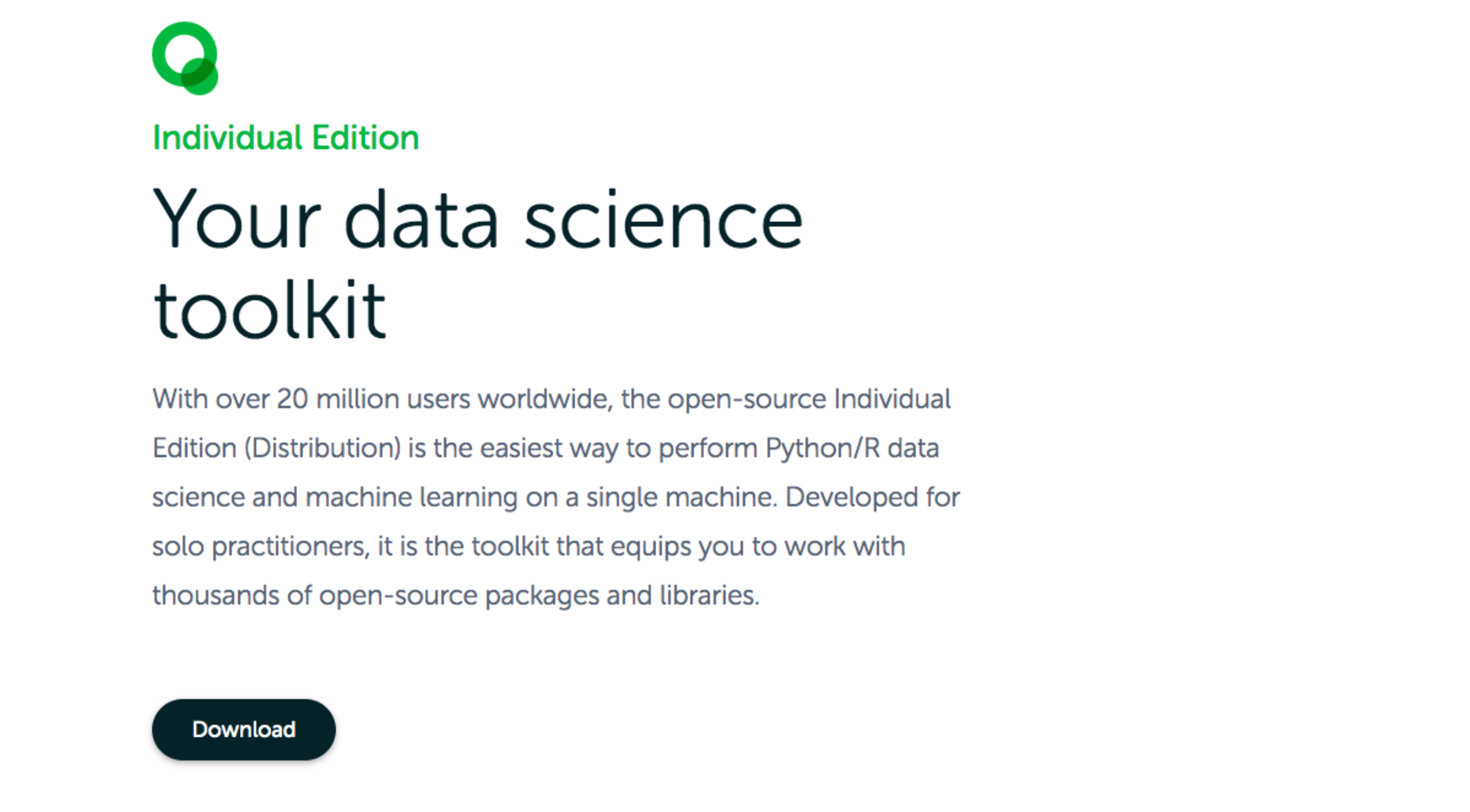
Get the Toolbox App to download PyCharm and its future updates with ease

```
[X]
  Cookies and IP addresses allow us to deliver and improve
  our web content and to provide you with a personalized
  experience. Our website uses cookies and collects your
  IP address for these purposes.

  =-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-
  |                                                      |
  | JetBrains may use cookies and my IP address to       |
  | collect individual statistics and to provide me with |
  | personalized offers and ads subject to the Privacy   |
  | Policy and the Terms of Use. JetBrains may use       |
  | third-party services for this purpose. I can revoke  |
  | my consent at any time by visiting the Opt-Out page. |
  |                                                      |
  | [Y]es, I agree    [N]o, thanks                       |
  |                                                      |
  =-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-

  ~ root#
```

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Popular IDE for Python: Anaconda

Anaconda installation site->
https://www.anaconda.com/products/individual



Individual Edition

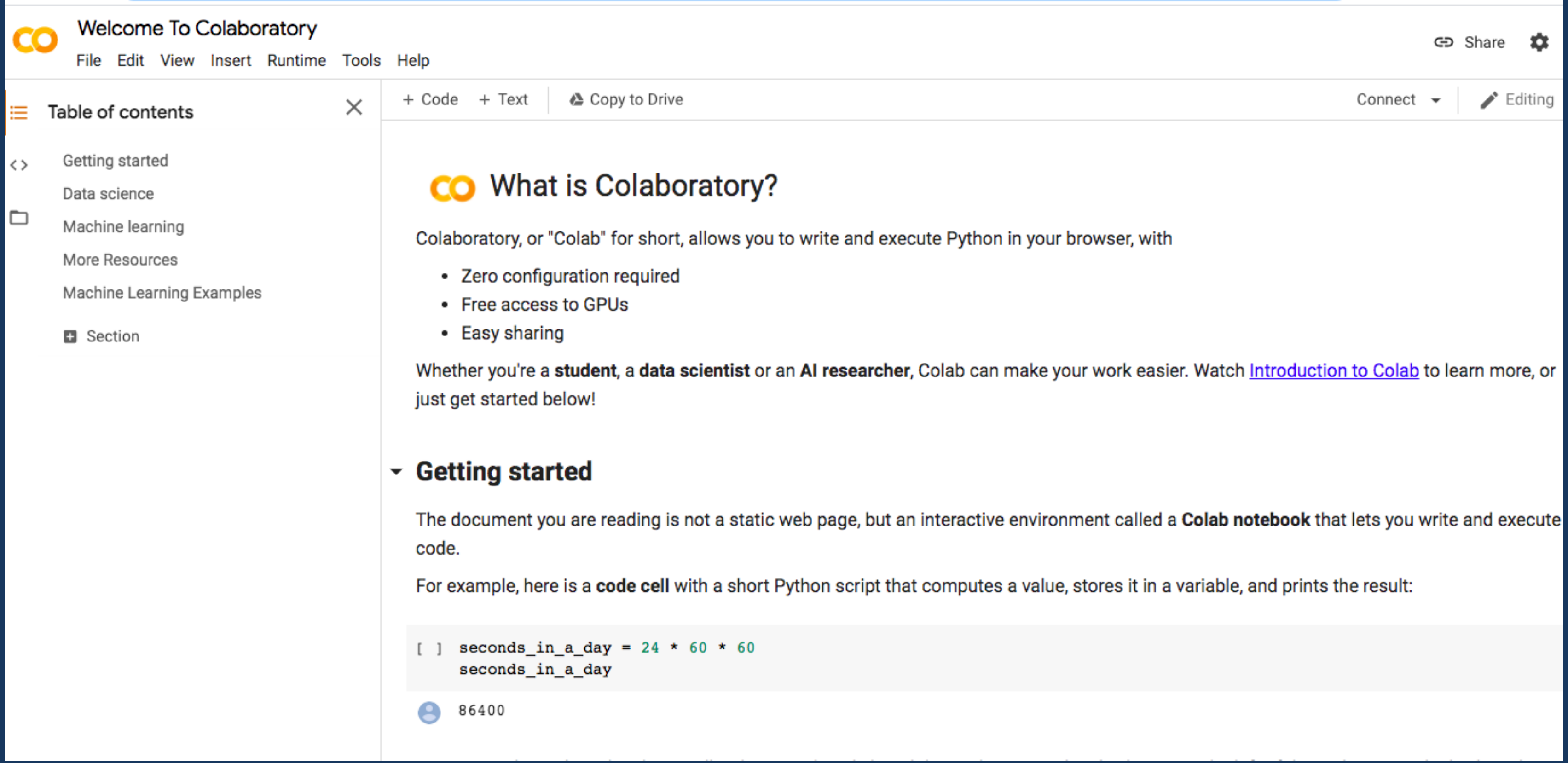
Your data science toolkit

With over 20 million users worldwide, the open-source Individual Edition (Distribution) is the easiest way to perform Python/R data science and machine learning on a single machine. Developed for solo practitioners, it is the toolkit that equips you to work with thousands of open-source packages and libraries.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Popular IDE for Python: Google colab

Google collaboratory link->
https://colab.research.google.com/notebooks/intro.ipynb



What is Colaboratory?

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser, with

- Zero configuration required
- Free access to GPUs
- Easy sharing

Whether you're a **student**, a **data scientist** or an **AI researcher**, Colab can make your work easier. Watch Introduction to Colab to learn more, or just get started below!

Getting started

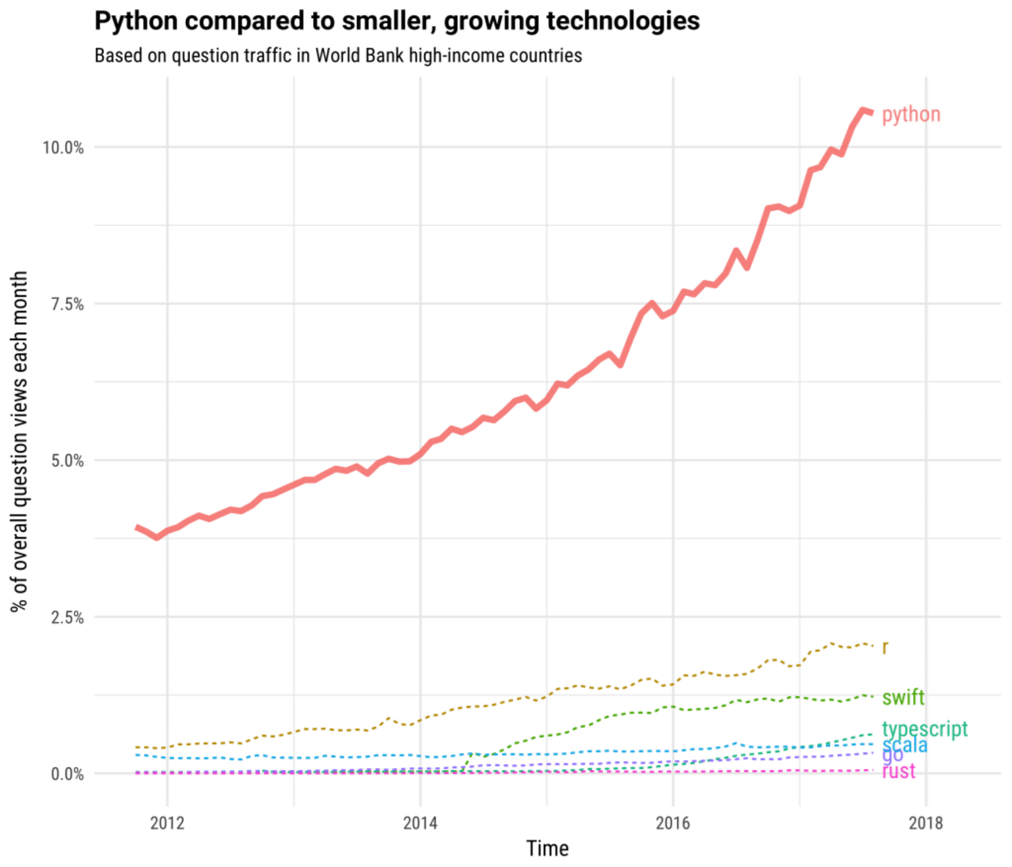
The document you are reading is not a static web page, but an interactive environment called a **Colab notebook** that lets you write and execute code.

For example, here is a **code cell** with a short Python script that computes a value, stores it in a variable, and prints the result:

```
[ ] seconds_in_a_day = 24 * 60 * 60
    seconds_in_a_day
```

```
86400
```

Statistical measurement on Python user

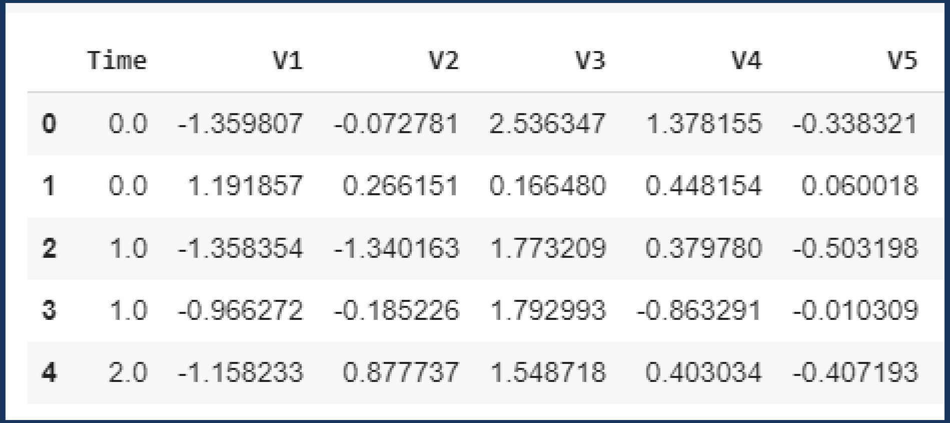


In recent time it is prominent that Python is one of the most popular language because of it’s simplicity

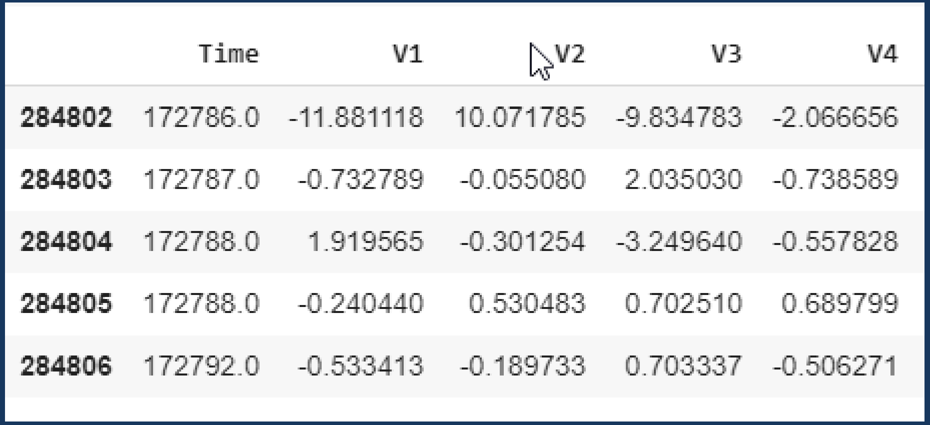
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

What is Machine Learning?

Machine learning is a sub-set of artificial intelligence (AI) that allows the system to automatically learn and improve from experience without being explicitly programmed



	Time	V1	V2	V3	V4	V5
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193



	Time	V1	V2	V3	V4
284802	172786.0	-11.881118	10.071785	-9.834783	-2.066656
284803	172787.0	-0.732789	-0.055080	2.035030	-0.738589
284804	172788.0	1.919565	-0.301254	-3.249640	-0.557828
284805	172788.0	-0.240440	0.530483	0.702510	0.689799
284806	172792.0	-0.533413	-0.189733	0.703337	-0.506271

Training Data

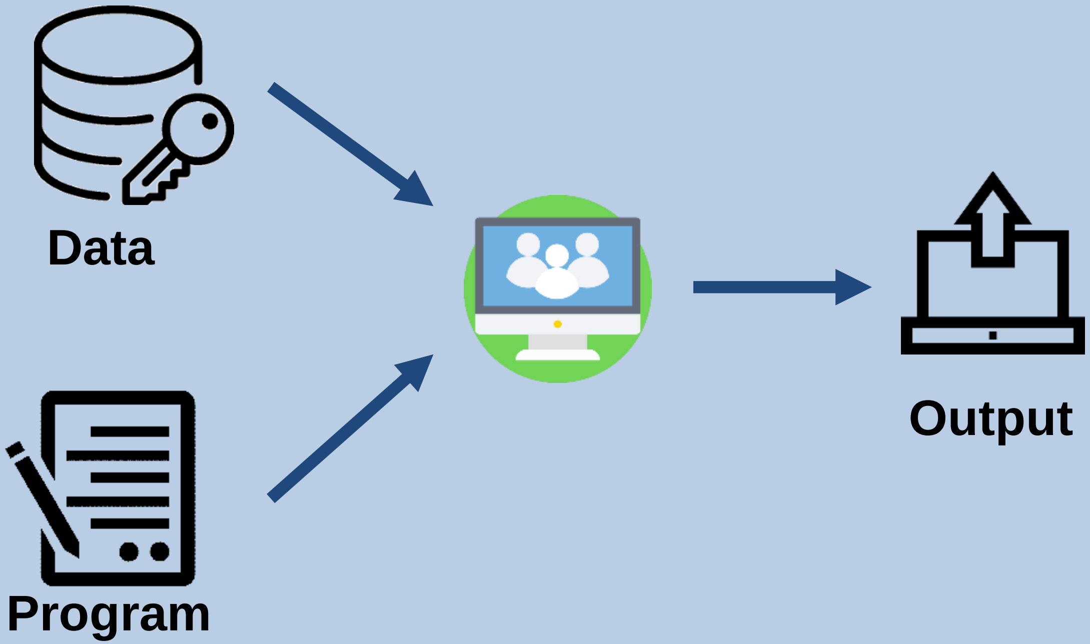
Model Building

Testing Data

Traditional Vs Machine Learning

greatlearning
Learning for Life

Traditional Programming

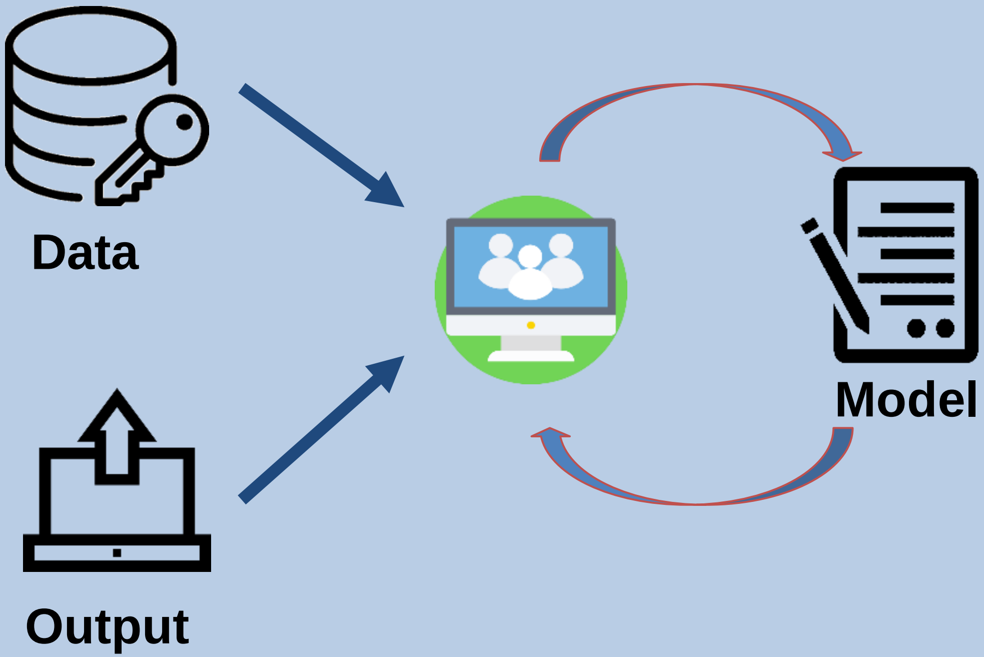


Data

Program

Output

Machine Learning

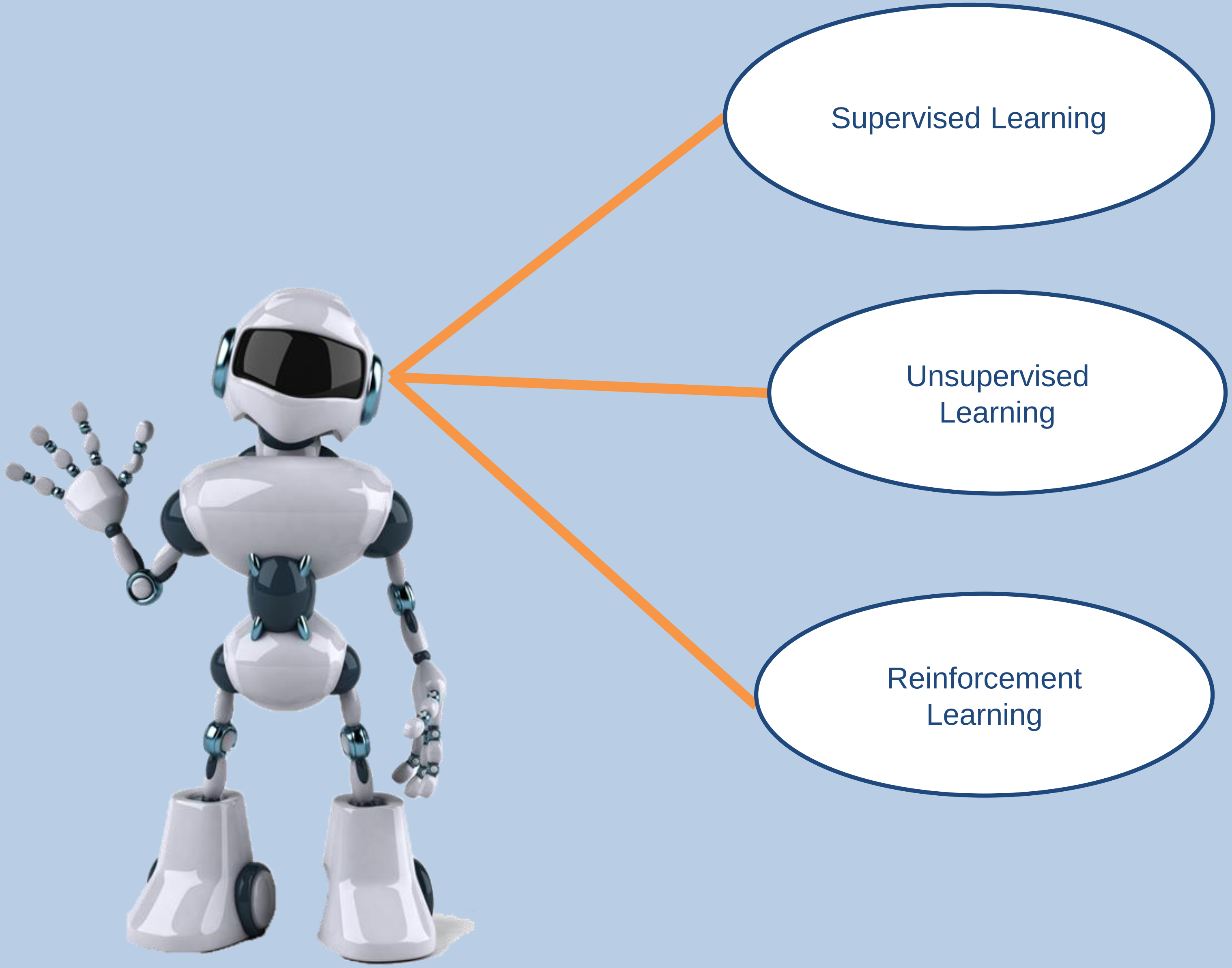


Data

Output

Model

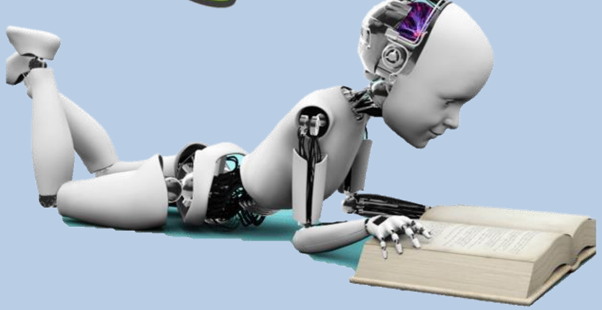
Types Of Machine Learning

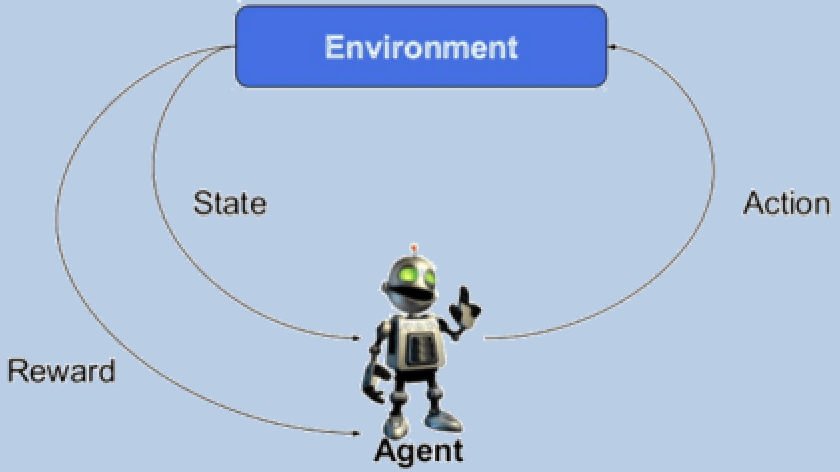


Supervised Learning

Unsupervised Learning

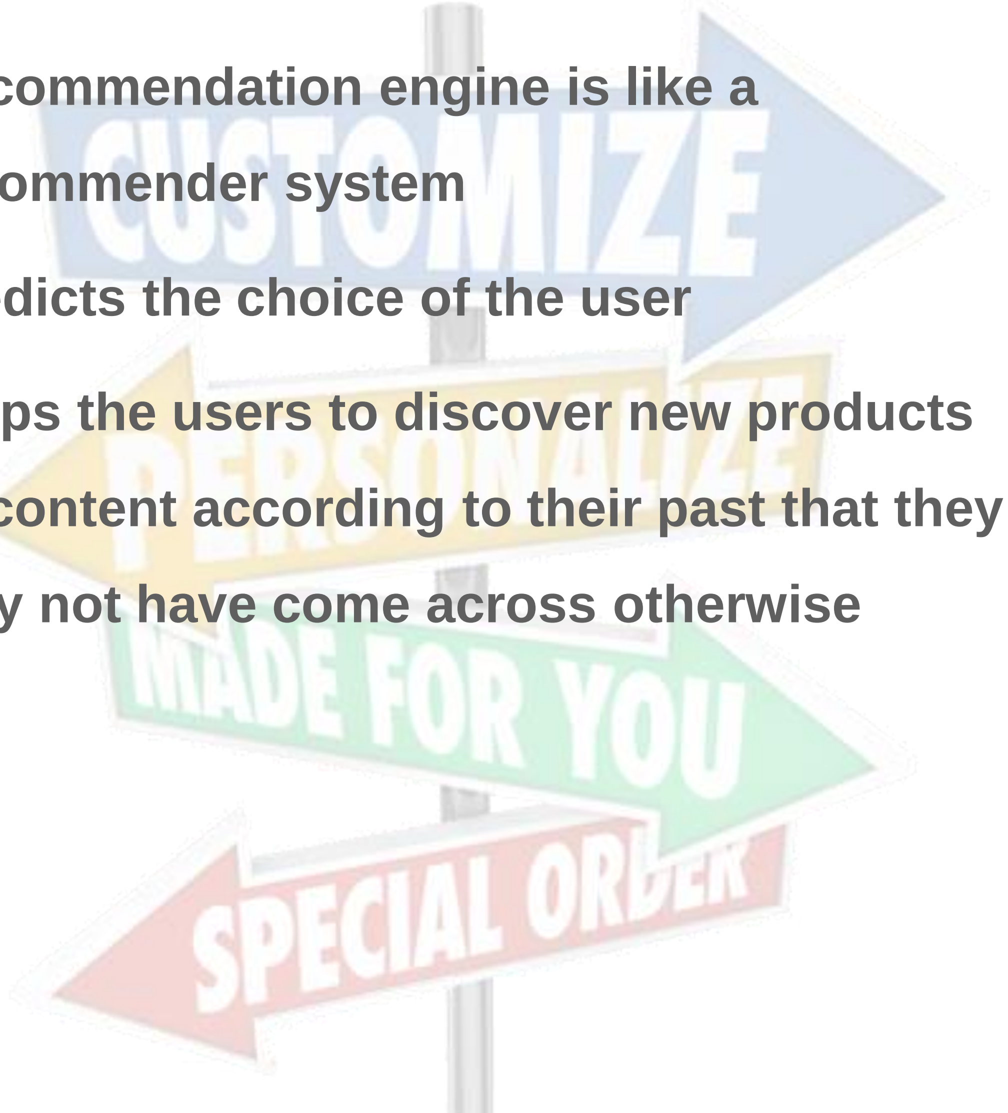
Reinforcement Learning





Recommendation Engine

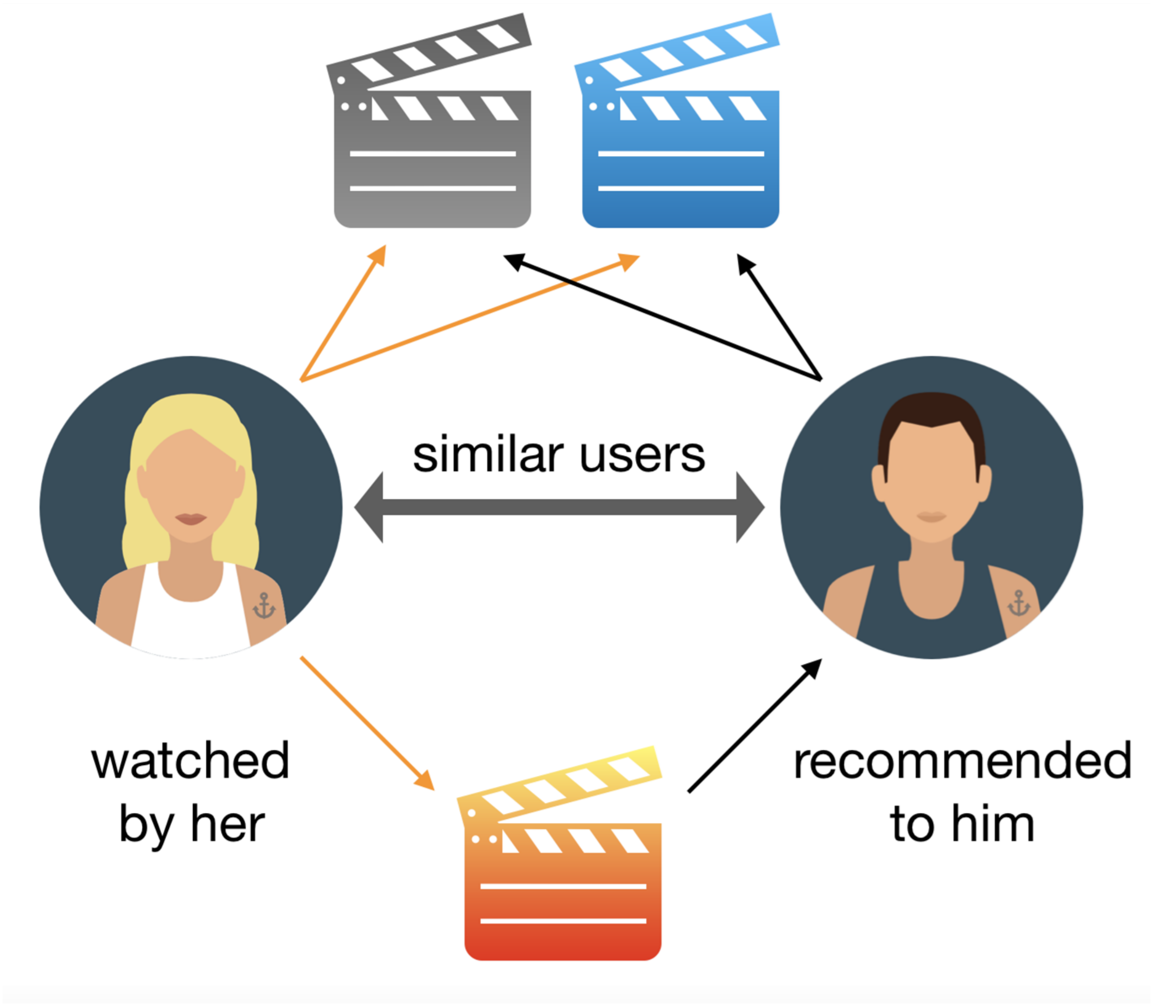
- Recommendation engine is like a recommender system
- Predicts the choice of the user
- Helps the users to discover new products or content according to their past that they may not have come across otherwise



CUSTOMERS WHO BOUGHT THIS ITEM:
ALSO BOUGHT:
piccolo

Product Recommendation

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited



Recommendation System

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Applications of Recommendation System

LinkedIn job matching algorithms has improved the performance by 50%

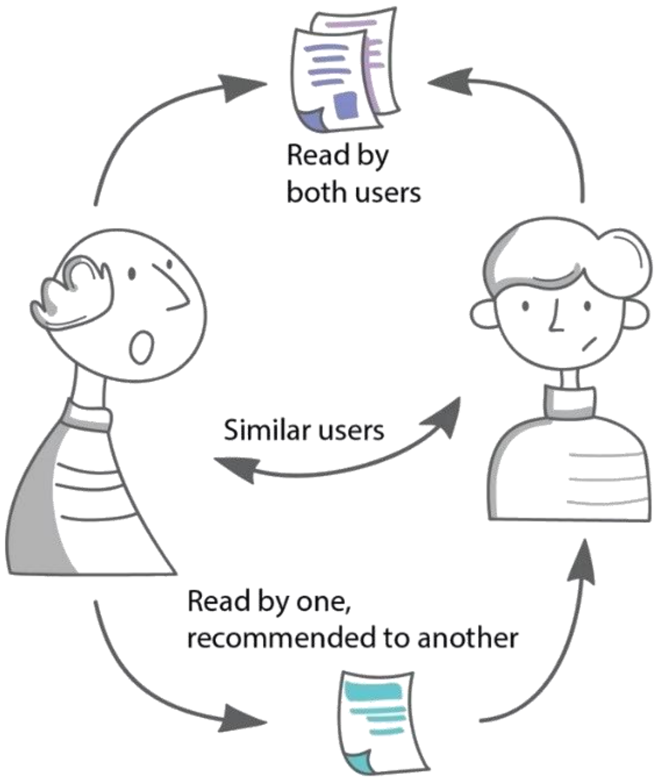
Netflix values recommendations at half a billion dollars to the company

Instagram switches to use algorithmic feed

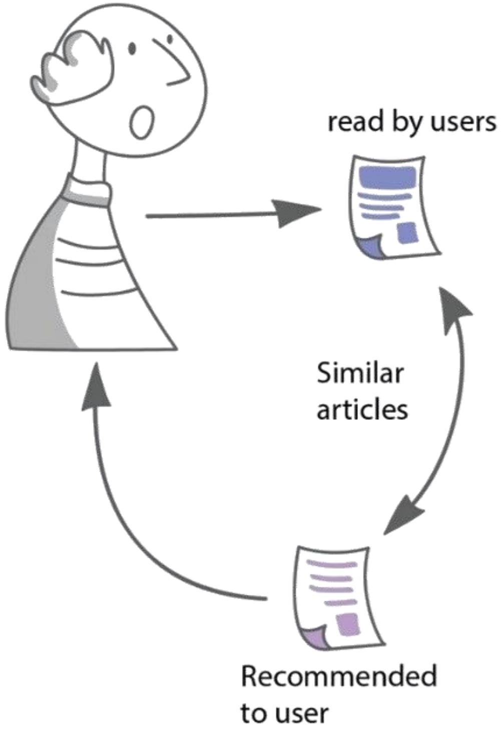
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Types of Recommendation System

User-based filtering



Content-based filtering

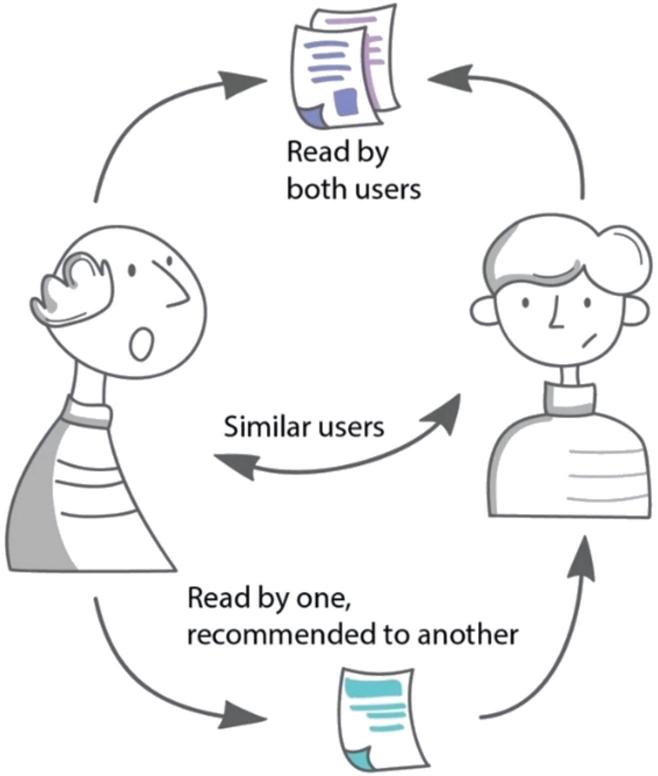


Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Types of Recommendation System: User-based

User-based filtering

Building a model from a user's past behavior as well as similar decisions made by other users. This model is then used to predict items that the user may have an interest in

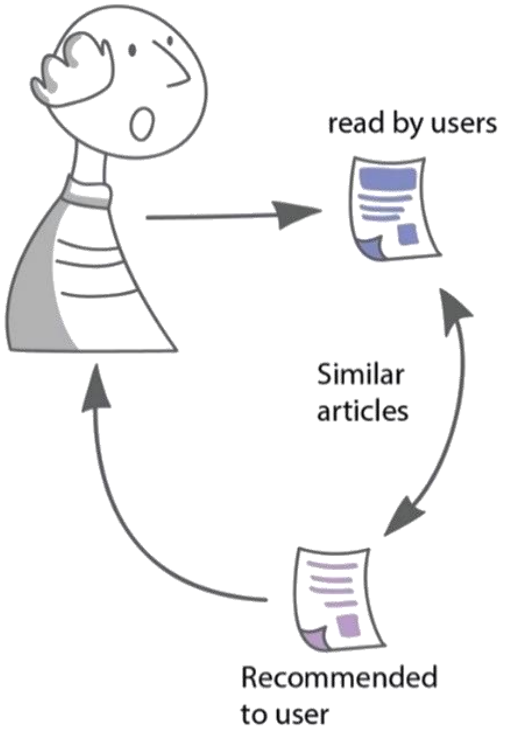


Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Types of Recommendation System: User-based

Content-based filtering

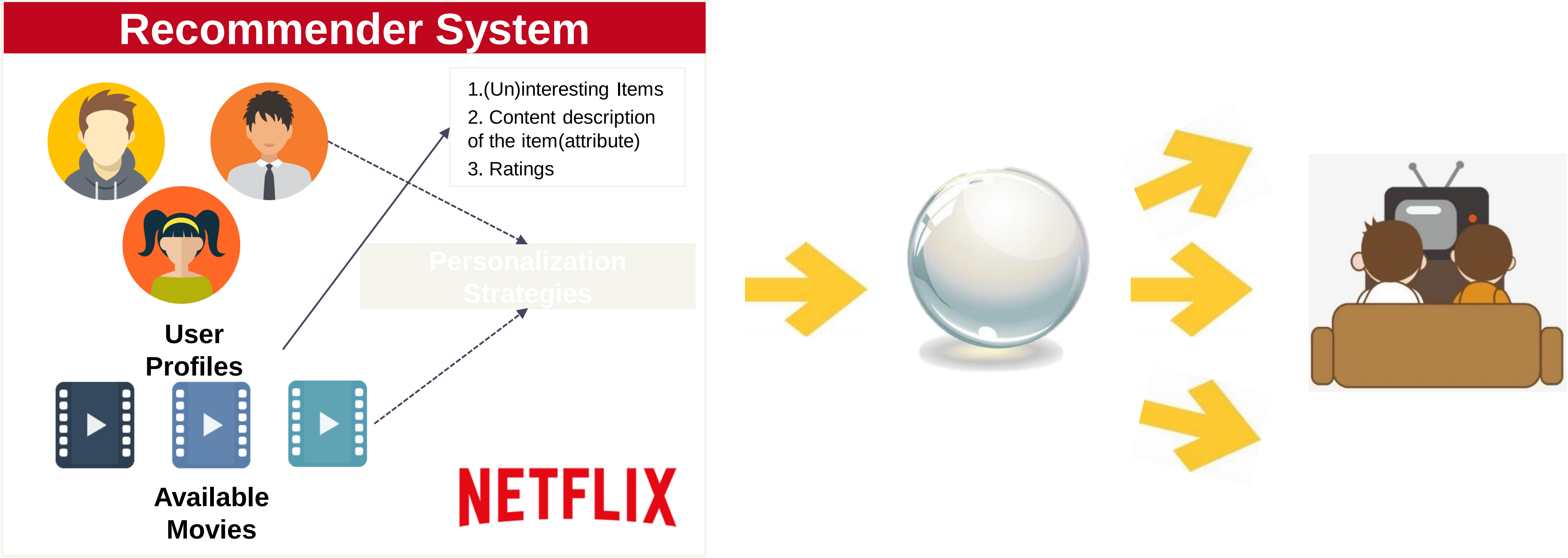
Utilizes a series of discrete characteristics of an item in order to recommend additional items with similar properties to the user



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Netflix: Hybrid Recommender System

A Hybrid recommender system is based on both the concept UBF and CBF



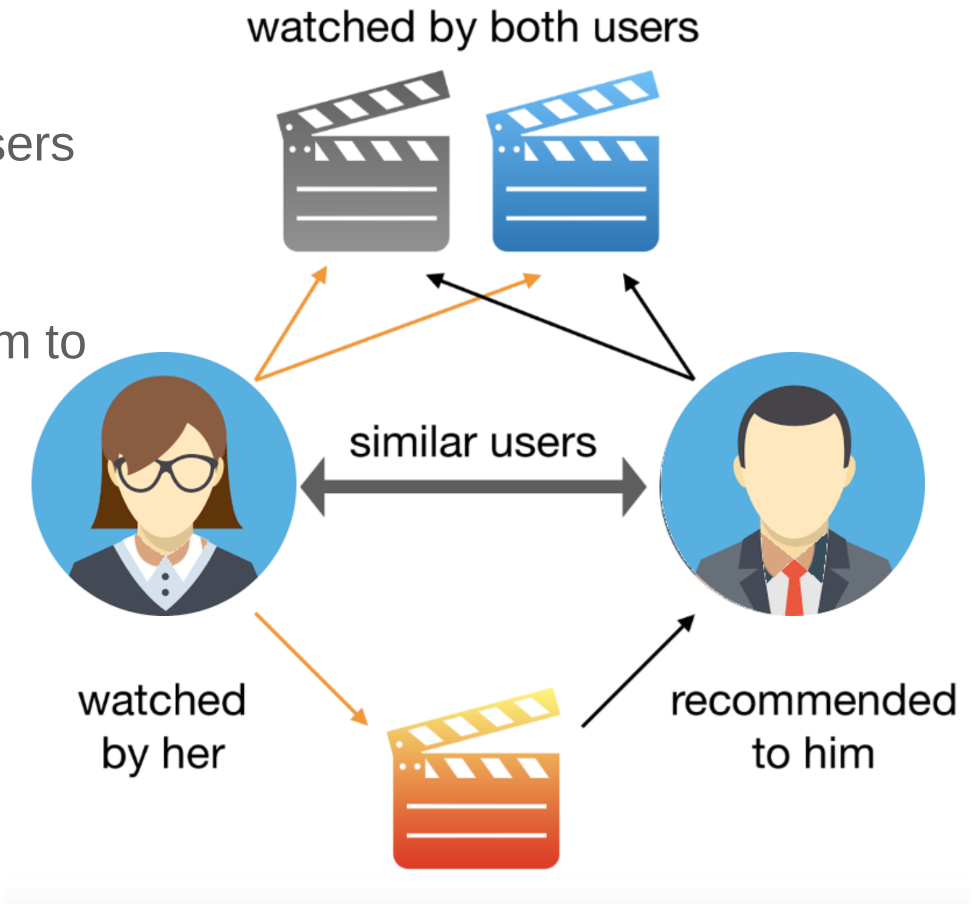
Recommender System

1.(Un)interesting Items
2. Content description of the item(attribute)
3. Ratings

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

User-Based Collaborative Filtering (UBCF)

- Algorithm finds the large group of users and also searches users with similar tastes
- Algorithm looks at different things they like and combines them to create a ranked list of suggestions
- **Algorithms used for measuring user or item similarity:**
  - **K – nearest neighbor (k-NN)**
  - **Pearson Correlation**



User-Based Collaborative Filtering (UBCF): Example

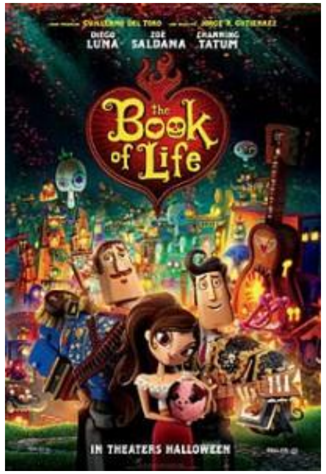
“Liza just watched the movie *Jocker*.”

Let’s see how the recommendation engine recommends her the movies?

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

User-Based Collaborative Filtering (UBCF): Example

Generate a list by the machine of users who have seen the following movies:



MARVEL
AVENGERS
AGE OF ULTRON
MAY 1

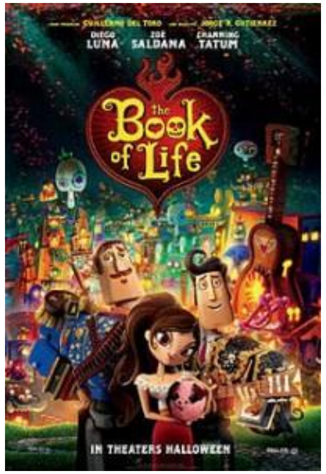
A NETFLIX ORIGINAL
A BLACK MIRROR EVENT
BANDERSNATCH
28 DEC
NETFLIX

Sam	Yes	Yes	Yes	Yes
Jay	No	Yes	No	No
Ratan	No	Yes	Yes	No
Dev	No	No	No	Yes

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

User-Based Collaborative Filtering (UBCF): Example

- **List of Liza's watched movies**
- **Find the same taste user**

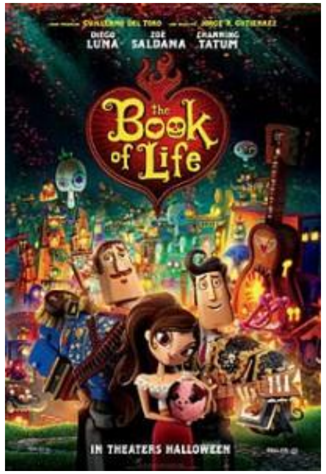


Sam	Yes	**Yes**	Yes	Yes
Jay	No	**Yes**	No	No
Ratan	No	**Yes**	Yes	No
Dev	No	No	No	Yes
Liza	?	Yes	?	?

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

User-Based Collaborative Filtering (UBCF): Example

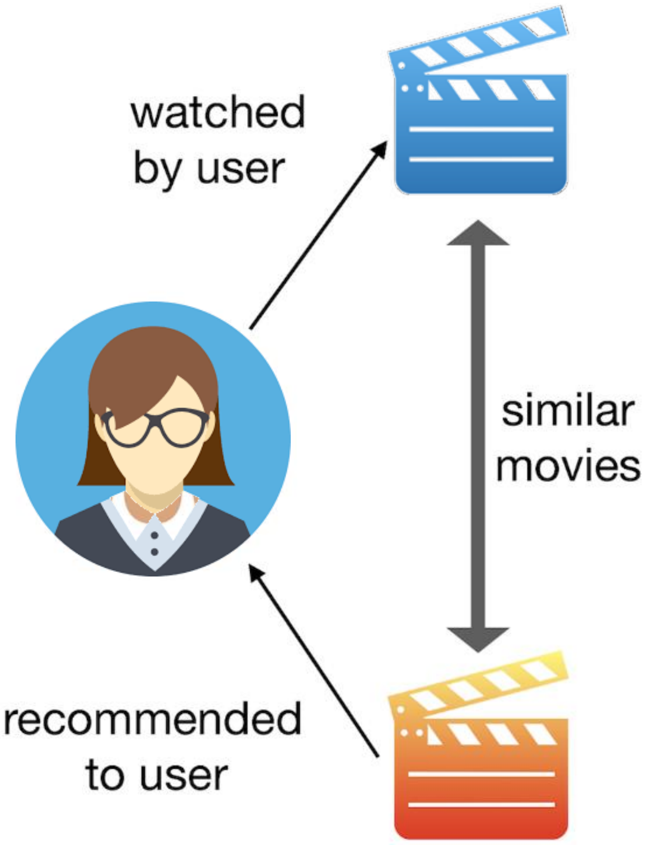
- List of Liza's watched movies
- Find the same taste user
- From the user's similarity, it's found the probable movie for Liza "Avengers" which gets more votes, so it gets recommended to Liza

				
Sam	Yes	**Yes**	Yes	Yes
Jay	No	**Yes**	No	No
Ratan	No	**Yes**	Yes	No
Dev	No	No	No	Yes
Liza	?	Yes	?	?
	1 vote		2 votes ✓	1 vote

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

User-Based Collaborative Filtering (UBCF)

- Contents are the central entities
- Works with data that the user provides, either *explicitly* (rating) or *implicitly* (clicking on a link, purchase history)
- Based on that data, the suggestions are given to the user
- Engine's accuracy increases with more input given to it



Content-Based Filtering (CBF): Example

“Liza just watched the movie *Jocker*.”

Let’s see how the recommendation engine recommends her the movies?

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

User-Based Collaborative Filtering (UBCF): Example

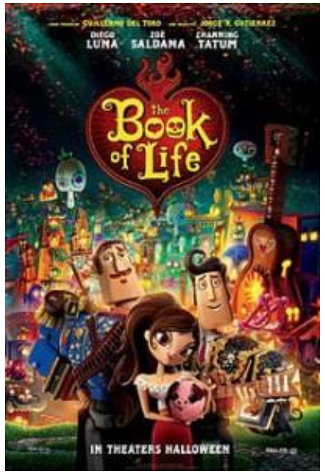
- **Generate a list of features about the movies like:**
  - **Actors**
  - **Directors**
  - **Themes**
  - **Story**
  - **Characters**

:

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

User-Based Collaborative Filtering (UBCF): Example

- Compare columns of each movies with column of the movie ***Jocker***



Hero	Yes ✓	Yes	No	Yes
Horror	No	No	No	Yes
Theme	Yes	Yes	Yes	Yes
IMDB rating 8+	Yes	No	Yes	No
Comedy	Yes	No	No	No

Most common movie is “The Book of Life”

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Demo: Movie Recommendation system

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Thank You

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited