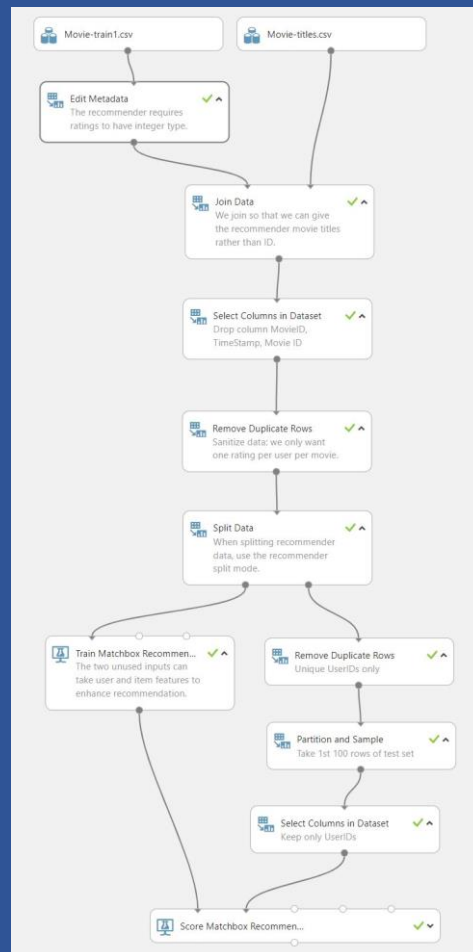# Movie

in Azure ML

In this session

- Question and Data
- Dataset description
- Create Data Set
- Place dataset
- Drop missing data rows
- Split Data
- Train, Score, Evaluate
- Understand Confusion matrix

GreatFriends.Biz

The finished model

https://raw.githubusercontent.com/laploy/ML.NET/master/Movie/movie-azure-model.JPG

Question and Data

Question: Which movie the user will also enjoy?

Dataset:

Train
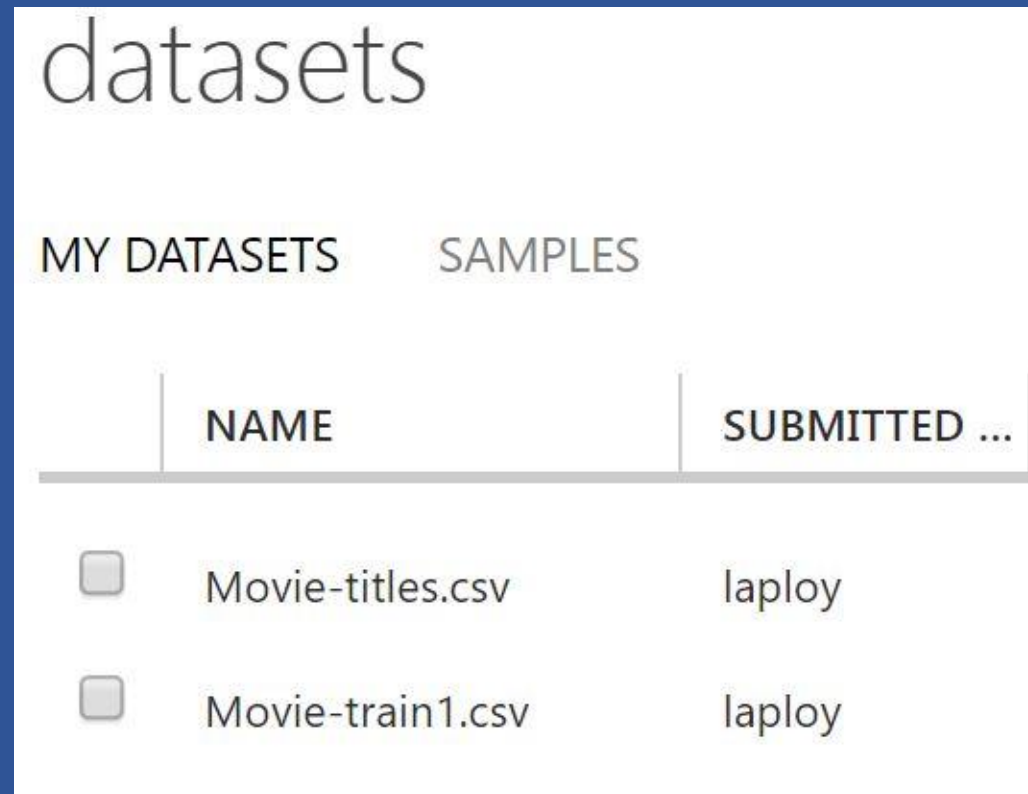https://raw.githubusercontent.com/laploy/ML.NET/master/Movie/Movie-train1.csv

Title
https://raw.githubusercontent.com/laploy/ML.NET/master/Movie/Movie-titles.csv

Dataset description

- User Id Feature
- moiveId Feature
- rating Label
- timestamp non-Feature

26,772 users
227,472 rows

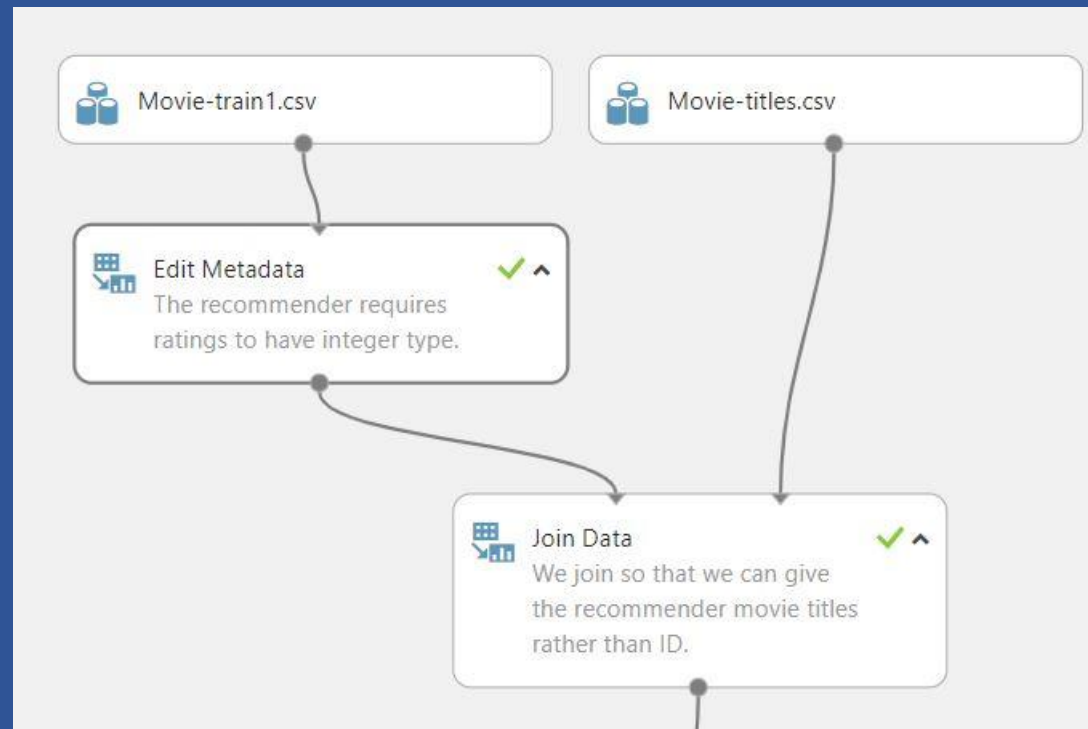|	A	B	C	D
1	UserId	MovieId	Rating	Timestamp
2	1	68646	10	1381620027
3	1	113277	10	1379466669
4	2	454876	8	1394818630
5	2	790636	7	1389963947
6	2	816711	8	1379963769
7	2	1091191	7	1391173869
8	2	1322269	7	1391529691
9	2	1433811	8	1380453043

Create Data Set

Click DATASET / + NEW / import both datasets
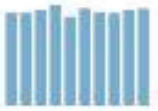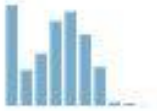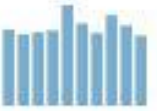
Add Datasets, Change rating to integer, Join

Data before join

Data after join

rows	columns				
227472	6				

UserId	MovieId	Rating	Timestamp	Movie ID	Movie Name
1	68646	10	1381620027	68646	The Godfather (1972)
1	113277	10	1379466669	113277	Heat (1995)
2	454876	8	1394818630	454876	Life of Pi (2012)
2	790636	7	1389963947	790636	Dallas Buyers Club (2013)
2	816711	8	1379963769	816711	World Z (2013)
2	1091191	7	1391173869	1091191	Lone Survivor (2013)
2	1322269	7	1391529691	1322269	August Osage County (2013)

view as

Drop column, remove duplicate rows, split
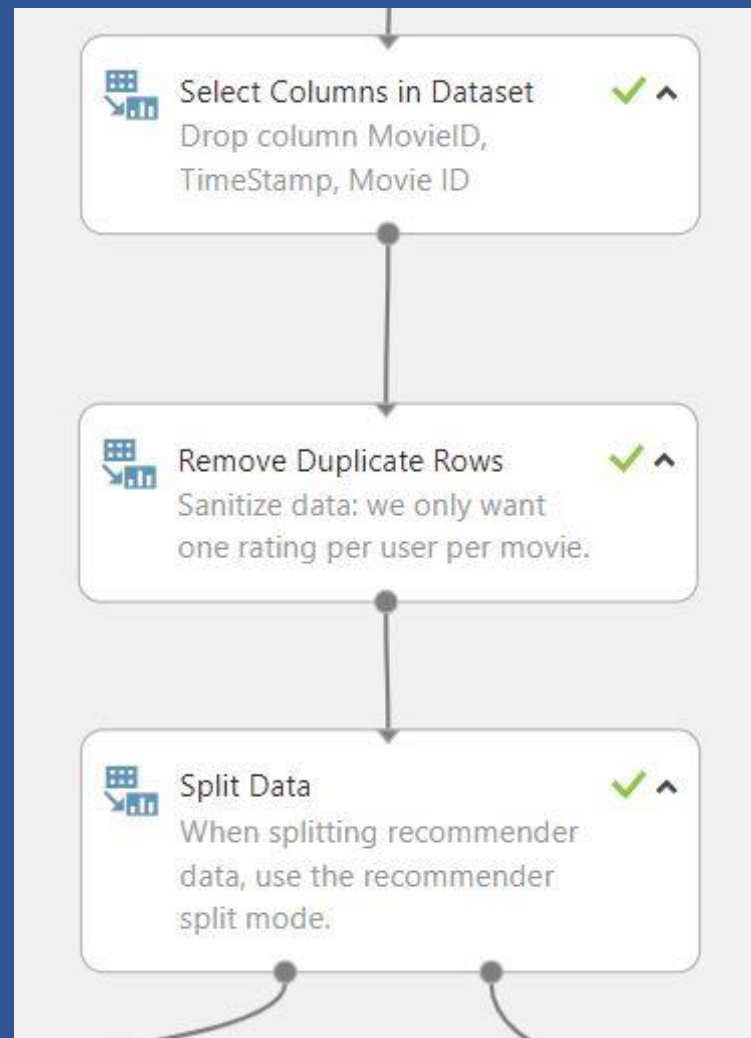
Score result

	User	Item 1	Item 2	Item 3
view as				
	12400	The Shawshank Redemption (1994)	The Godfather (1972)	The Dark Knight (2008)
	587	The Shawshank Redemption (1994)	The Godfather (1972)	The Godfather Part II (1974)
	14887	The Shawshank Redemption (1994)	The Godfather (1972)	The Dark Knight (2008)
	19075	The Shawshank Redemption (1994)	The Godfather (1972)	The Dark Knight (2008)

What next?

Write code for Movie Recommendation