Introduction to DMC 2022

The Task

- Prediction of replenishments of products per user
  - One prediction per user-product combination
  - Prediction: either 0 (no replenishment) or 1-4 (week of replenishment)
  - 3 points if replenishment week is predicted correctly,
    1 point if replenishment or no replenishment is predicted correctly
- Data: *items.csv, orders.csv, category_hierarchy.csv, submission.csv*
- Time frames
  - Training: 01.06.2020 – 31.01.2021
  - Prediction: 01.02.2021 – 28.02.2021
  - Internal Prediction Periods: 01.12.2020 – 28.12.2020 and
    01.01.2021 - 28.01.2021
- Files for internal prediction periods in ILIAS

items

itemID	brand	feature_1	feature_2	feature_3	feature_4	feature_5	categories
22665	861	4	0	490	2	66	[2890, 855, 3908, 3909]
28640	1366	10	1	537	0	101	
13526	1090	10	0	511	0	0	[3270, 163, 284, 1694, 12, 3837, 2422, 3595, 3561, 3912, 3913, 3914, 3915]
21399	1090	10	1	511	0	0	[3270]
8504	768	4	1	484	0	66	[2470]
32122	5	4	1	491	0	66	
31956	1388	4	0	491	0	66	[236, 3625, 356]
6237	1492	4	1	491	3	66	[2658, 1686]
16971	288	6	0	314	0	45	[390, 2678, 1708]
18385	288	6	0	314	0	45	[390, 1708]
12775	322	1	0	536	0	78	[3222, 1170, 2296, 3751, 3399, 1772, 3917]
8285	734	10	0	510	3	132	[3259, 2144, 2984, 677, 3349]
16238	1180	10	0	288	0	87	[2942]
29828	1180	10	0	288	0	87	[2942]
13025	589	10	1	503	3	17	[1091, 2325]

University of Mannheim
School of Business Informatics and Mathematics

orders

date	userID	itemID	order
2020-06-05	0	20664	1
2020-08-03	0	15083	1
2020-08-03	0	12468	1
2020-08-18	0	12505	1
2020-09-01	0	1505	1
2020-10-09	0	26387	1
2020-10-09	0	20664	1
2020-11-20	0	9325	1
2020-11-20	0	28231	1
2020-12-04	0	31683	1
2020-12-04	0	31923	1
2020-12-11	0	28231	1
2020-12-11	0	6446	1
2020-12-11	0	20664	1
2021-01-15	0	6446	1

category_hierarchy

category	parent_category
0	75
1	1499
2	1082
3	3498
4	1623
5	2478
6	1582
7	3027
8	2364
9	3590
10	582
11	3686
12	3241
13	510
14	2160

submission

userID	itemID	prediction
0	20664	
0	28231	
13	2690	
15	1299	
15	20968	
20	8272	
24	11340	
34	21146	
34	31244	
46	31083	
61	4648	
76	4603	
76	23869	
76	29829	
78	12635	

TOPICS
- Orders
- Categories
- Seasonality
- Train/Validation/Test

The Evaluation

```python
def count_points(pred, gold):
    df = pd.merge(pred, gold, on=['userID', 'itemID'], suffixes=('_pred', '_gold'))
    df['points'] = df.apply(_compute_points_for_row, axis=1)
    return df['points'].sum()


def _compute_points_for_row(row):
    y_pred, y_gold = row.prediction_pred, row.prediction_gold
    if y_pred == y_gold:
        # one point if "no order" (0) is predicted correctly; three points if order week is predicted correctly
        return 1 if y_pred == 0 else 3
    # one point if order is predicted correctly (but not the correct week), otherwise zero points
    return 1 if (y_pred > 0 and y_gold > 0) else 0
```

BASELINES

1) Always predict „1"

2) Predict based on previous month

See ILIAS for splitted datasets and evaluation code (check!)

The Leaderboard

- Go to http://ilias2.informatik.uni-mannheim.de

- Log in with your team login and passwort "DMC2022"
  - Change your password after first login via Settings

- Click: "REPOSITORY" -> "Repository – Home"
  -> "Data Mining Cup 2022" -> "Start the Test"

- Submit your files and have a look at the ranking in "Results"

- Submit your best solution once a week before our meeting!

Q & A

Your Task for 03/05/20

- Every team presents roughly 10 minutes about
  - Analyses, insights, questions concering the data
  - Their first approach of predicting replenishment
    - don't forget to include the performance on the test sets!
  - Anything else they find interesting about the task
    - Problems they encountered (and solved)
    - Unclear details of the task
    - …