# Recommendation System

Internship
Festo Owiny, festo.owiny@studenti.unitn.it

## I. Introduction & Motivation

Mobile communication systems have experienced fast development over the years with the recent development of 5G LTE systems. LTE systems have facilitated high data rate, reduced latency, cost reduction, and massive device connectivity. In turn, mobile Internet and e-commerce have experienced fast development. With the development of e-commerce technology, a growing number of people prefer to purchase items on the e-commerce websites [1].

Online stores have become a very popular choice for many customers because it usually provides them with a large number of products. However, often customers can not accurately find their favorite products and they waste considerable amount of time browsing. Therefore, it is inevitably necessary for online stores to recommend appropriate items to their customers. An effective Recommendation System (*RS*) is necessary for customers. This need is also strongly required for *3A*.

*RS* learns from data and recommends most valuable products to a customer. RS is therefore a filter system that selects a few products, helping users find relevant items within a huge list. This enables personalization resulting into better customer satisfaction and reduced churn. User-based Collaborative Filtering (UCF) algorithm is widely utilized to predict the preferences of customers.

Recent statistics from *Econsultancy* and *Monetate* report 94% of companies agree that personalization is critical to the current and future successes [3]. The system is currently used in Amazon, eBay, Netflix, Facebook and others. Amazon's RS drives a 20-35% lift in sales annually [4].

## II. Dataset

Dataset is extracted from the database between 2017 and 2018, excluding customers whose names start with *T*. Data is presented in the (*.csv*) format. An *sql* script was embedded within the *Python* script for direct data retrieval from the database. The data represents a bunch of nested *sql* queries within the company database. The major feature used from our data in the proposed algorithm is the customer purchase history.

## III. Marketing and sales strategy

**Up-sell**: Persuade a customer to buy more (of the same product type) than the customer had originally intended.

- Sell a more expensive later model of a product.
- Sell multiple units: Buy 2 get 1 free.
- Add on a warranty & insurance.

**Cross-sell**: Persuade a customer to buy more of the other products than the customer had originally intended.

- Improvement in customer satisfaction score and customer lifetime value (reduced Churn).

## IV. RS Model

### A. Collaborative Filtering

The method uses subset of users with similar tastes to recommend items for a target user. The assumption is that; Users with similar interests have common preferences (item to item correlation). The advantage of this method is that no knowledge about item features is needed. It provides better scalability because of correlations between a limited number of items instead of using very large number of product features. There is also reduced sparsity problem in the matrix which is a common problem in recommendation systems. The key problem is the cold start problem whereby a client has no purchase history. Here, we recommend with other techniques such as Popularity-based.

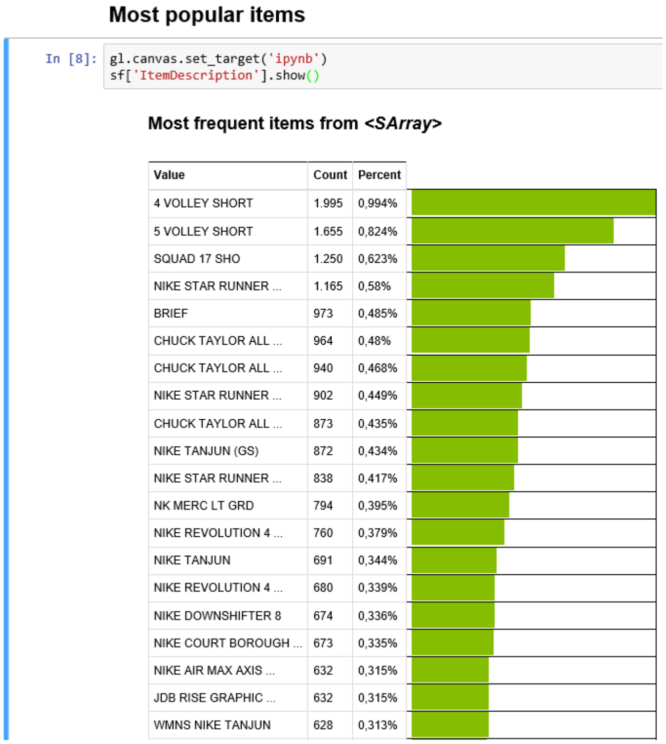
Basic Algorithms used in collaborative filtering matches similar items based on *Jaccard* and *Cosine* Similarity [1]. We represent different customers as vectors. Each dimensionality of the vector represents the purchasing frequency of an item. Purchasing frequency indicates the counts of items purchased by a user. We consider the purchasing frequency of different users. Appropriate arithmetic techniques corresponding to Jaccard or Cosine Similarity is then employed between different customer pairs.

### B. Popularity-based recommender

Popularity defines the degree of success of an item in its market. Popularity feedback can be a powerful determinant of online behavior. Examples are; most popular fashion, music downloads, movie ratings and number of views. Popularity is adopted whenever the cold start problem is encountered, that is when a client has no purchase history. The technique can also adopted to suggest most popular items to a client. With regard to *3A*, Popular purchase pairs can be suggested to both new and repeat clients. The limitation of Popularity-based recommender is lack of personalization.

## V. Implementation

We used both the personalized model (Collaborative Filtering) and the Popularity-based model to perform our recommendation. The project was implemented in Python with associated data science libraries such as Graphlab, Pandas, Numpy, and others. The tools were also used for data visualization. A



Most popular items

```
In [8]: gl.canvas.set_target('ipynb')
        sf['ItemDescription'].show()
```

Most frequent items from <SArray>

| Value | Count | Percent |
|---|---|---|
| 4 VOLLEY SHORT | 1.995 | 0,994% |
| 5 VOLLEY SHORT | 1.655 | 0,824% |
| SQUAD 17 SHO | 1.250 | 0,623% |
| NIKE STAR RUNNER ... | 1.165 | 0,58% |
| BRIEF | 973 | 0,485% |
| CHUCK TAYLOR ALL ... | 964 | 0,48% |
| CHUCK TAYLOR ALL ... | 940 | 0,468% |
| NIKE STAR RUNNER ... | 902 | 0,449% |
| CHUCK TAYLOR ALL ... | 873 | 0,435% |
| NIKE TANJUN (GS) | 872 | 0,434% |
| NIKE STAR RUNNER ... | 838 | 0,417% |
| NK MERC LT GRD | 794 | 0,395% |
| NIKE REVOLUTION 4 ... | 760 | 0,379% |
| NIKE TANJUN | 691 | 0,344% |
| NIKE REVOLUTION 4 ... | 680 | 0,339% |
| NIKE DOWNSHIFTER 8 | 674 | 0,336% |
| NIKE COURT BOROUGH ... | 673 | 0,335% |
| NIKE AIR MAX AXIS ... | 632 | 0,315% |
| JDB RISE GRAPHIC ... | 632 | 0,315% |
| WMNS NIKE TANJUN | 628 | 0,313% |

Figure 1. Popular items

description of different scripts used for our results are included in the *ReadMe.txt* file.

I then used *SSH* software package to enable secure system administration and file transfer for access of my results by my colleagues as required. I also wrote a python script to perform this action.

*.bat* files were written for the associated python scripts for simpler execution. This also helped to hide complexity of the individual ordered running of every scripts, because a single run of *run.bat* was able to run all the associated python scripts. The resulting recommendations for each client are saved in *.json* and *.csv* file formats as required. The files associated with the recommendation system is described in the attached *ReadMe.txt* file.

The figures in Section VI illustrate some of the analyses printed.

## VI. Results

Figure 1 is the result for the most popular items. This list is the same for a given dataset. The most popular items can be recommended to new customers or other customers who might be interested as guided by the marketing department.

Figure 2 is the list of top 10 items recommended for the client *Gonzato* using collaborative filtering. Separate recommendation lists were provided for all the customers. We obtained an array of weights for items and arranged them in decreasing order. Then we selected top $N$ items as recommendations for a target user.

Figure 3 lists items most similar to *Nike Star Runner (TDV)*. The customers who purchased *Nike Star Runner (TDV)* also might have bought other products. We therefore retrieve a list of items most commonly bought together with *Nike Star Runner (TDV)*. We select top $N$ items as our most similar items to the target item. This analysis is performed for all or some items as guided by the marketing team.

Recommend products for "GONZATO SRL"

```
In [29]: pers_model.recommend(['GONZATO SRL'])
```

Out[29]:

| CustomerName | ItemDescription | score | rank |
|---|---|---|---|
| GONZATO SRL | M NSW NIKE AIR PANT FLC | 0.0307683111893 | 1 |
| GONZATO SRL | M NSW HBR+ HOODIE FZ FLC | 0.0306539005703 | 2 |
| GONZATO SRL | M NSW NIKE AIR HOODIE FZ FLC ... | 0.0298883815606 | 3 |
| GONZATO SRL | NIKE COURT BOROUGH LOW (TDV) ... | 0.0284595201413 | 4 |
| GONZATO SRL | WMNS NIKE COURT ROYALE AC | 0.0283351295524 | 5 |
| GONZATO SRL | Y NK GMSK - GFX | 0.0278258124987 | 6 |
| GONZATO SRL | M NSW NIKE AIR CREW FLC | 0.0268786847591 | 7 |
| GONZATO SRL | M NSW HBR+ JGGR | 0.0237348520093 | 8 |
| GONZATO SRL | M NSW HBR HOODIE PO FLC JDI ... | 0.0237329214811 | 9 |
| GONZATO SRL | M NSW TEE BRAND MARK | 0.0231589327256 | 10 |

[10 rows x 4 columns]

Figure 2. Top 10 items recommended for Mr. Gonzato

Items purchased with "NIKE STAR RUNNER (TDV)"

```
In [53]: pers_model.get_similar_items(['NIKE STAR RUNNER (TDV)'])
```

Out[53]:

| ItemDescription | similar | score | rank |
|---|---|---|---|
| NIKE STAR RUNNER (TDV) | NIKE STAR RUNNER (PSV) | 0.653645813465 | 1 |
| NIKE STAR RUNNER (TDV) | NIKE STAR RUNNER (GS) | 0.551999986172 | 2 |
| NIKE STAR RUNNER (TDV) | NIKE REVOLUTION 4 (TDV) | 0.399999976158 | 3 |
| NIKE STAR RUNNER (TDV) | NIKE COURT BOROUGH MID (PSV) ... | 0.398395717144 | 4 |
| NIKE STAR RUNNER (TDV) | NIKE COURT BOROUGH MID (GS) ... | 0.397619068623 | 5 |
| NIKE STAR RUNNER (TDV) | NIKE REVOLUTION 4 (PSV) | 0.362244904041 | 6 |
| NIKE STAR RUNNER (TDV) | NIKE AIR MAX AXIS (GS) | 0.334134638309 | 7 |
| NIKE STAR RUNNER (TDV) | NIKE COURT BOROUGH LOW (GS) ... | 0.331695318222 | 8 |
| NIKE STAR RUNNER (TDV) | NIKE COURT BOROUGH LOW (PSV) ... | 0.317333340645 | 9 |
| NIKE STAR RUNNER (TDV) | NIKE REVOLUTION 4 EU | 0.313868641853 | 10 |

[10 rows x 4 columns]

Figure 3. Top 10 items purchased with Nike Star Runner (TDV)

## VII. Conclusion

Recommendation system is widely used in e-commerce to assist customers to find their products of interest. This paper implements the popular Collaborative Filtering recommendation system and discusses its application with respect to *3A*. The system presents a better similarity among users and products. Therefore, it can make better recommendation to the customers. However, a hybrid of collaborative-based filtering, content-based filtering and other recommendation techniques can be employed for better efficiency. We could also perform customer clustering, then implement recommendations within the clusters.

## References


[1] Lexi Xu, Yue Chen *Clothing Recommendation System Based on Advanced User-Based Collaborative Filtering Algorithm*. , 2018.

[2] Liu, Xu, Chen, Fan *A novel power control mechanism based on interference estimation in LTE cellular networks*. , 2016.

[3] David Moth , Econsultancy *https://econsultancy.com/94-of-businesses-say-personalisation-is-critical-to-their-success/*. , 2013.

[4] http://rejoiner.com/resources/amazon-recommendations-secret-selling-online/ *https:The Amazon Recommendations Secret to Selling More Online*. , 2016.