DRIVES

Development and Research on Innovative Vocational Education Skills

U2 MACHINE LEARNING OVERVIEW

U2.E2 MACHINE LEARNING AND RECOMMENDATION SYSTEMS

Machine Learning Engineer

January 2021, Version 1

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

LEARNING OBJECTIVES

DRIVES
Development and Research on Innovative Vocational Education Skills

The student is able to

MLE.U2.E2.PC1	Know how to define recommendation systems.
MLE.U2.E2.PC2	Understand the operation and logic behind recommendation systems.
MLE.U2.E2.PC3	Understand the role of machine learning in recommendation systems.

TARGET

Recommender systems are a classic example of machine learning applications. The aim of the recommendation system is to make meaningful recommendations to users. These recommendations may include products, investments, services, etc. Suggestions for Amazon books or Netflix films are real examples of how industry recommendation systems operate.

LOGIC

The data required for the recommendation systems derive from explicit user ratings, behavior and preferences using search engines and purchase history or other knowledge of users and items themselves. Recommender systems use these insights and information about other users to determine what may be relevant to the user by using data mining techniques along with prediction algorithms.

RECOMMENDATION SYSTEMS: WHY DO WE NEED THEM?

Companies using recommendation systems focus on increasing sales as a result of highly customized offerings and enhanced customer experience.

Recommendations usually speed up searches and make it easier for users to access the content of their interest and to surprise them with offers they would never have been looking for.

Companies can win and retain customers by sending e-mails with links to new offers that meet the interests of their recipients or suggestions for products that fit their profile.

RECOMMENDATION SYSTEMS: WHY DO WE NEED THEM?

The user begins to feel known and understood and is more likely to purchase additional products or consume more content.

By knowing what the user wants, the company gains a competitive advantage and the risk of losing the customer to a rival company decreases.

The provision of added value to users is appealing, allowing companies to position themselves ahead of their competitors and eventually increase their earnings.

VALUE OF A RECOMMENDATION SYSTEM

Value for the customer

–Find things that are interesting

–Narrow down the set of choices

–Help me explore the space of options

–Discover new things

–Entertainment

…

Value for the provider

–Additional and probably unique personalized service for the customer

–Increase trust and customer loyalty

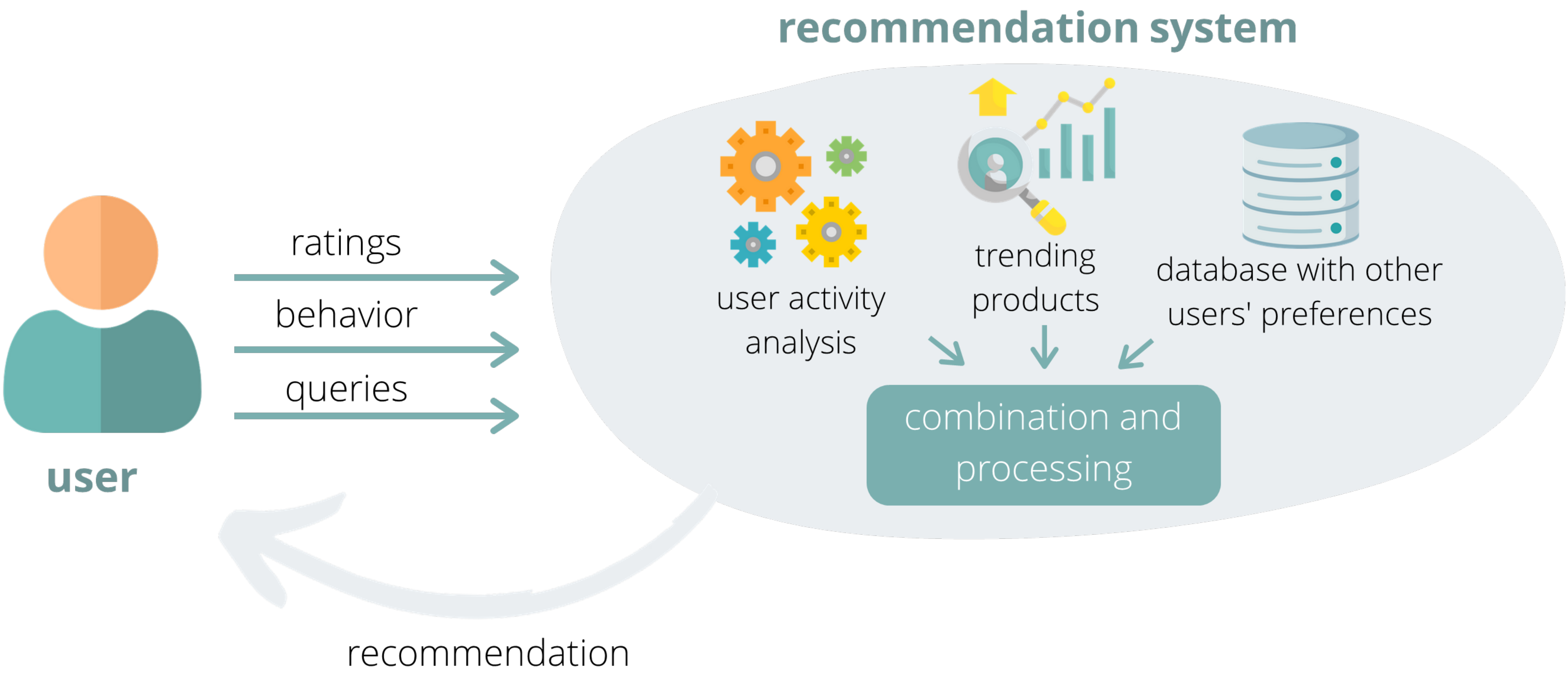
–Increase sales, click trough rates, conversion etc.

–Opportunities for promotion, persuasion

–Obtain more knowledge about customers

…

RECOMMENDATION SYSTEMS: HOW DOES IT WORK?



RECOMMENDATION SYSTEM SEEN AS A FUNCTION

DRIVES
Development and Research on Innovative Vocational Education Skills

Given:

–User model (e.g. ratings, preferences, demographics, situational context)

–Items (with or without description of item characteristics)

Find:

–Relevance score. Used for ranking.

Finally:

–Recommend items that are assumed to be relevant

But:

–Remember that relevance might be context-dependent

–Characteristics of the list itself might be important (diversity)

RECOMMENDATION SYSTEMS: HOW DOES IT WORK?

Recommendation systems work with two types of information:

CHARACTERISTIC INFORMATION

Information on items, such as keywords and categories, and information about users, such as preferences and profiles.

ITEM-USER INTERACTIONS

Information such as ratings, number of purchases, likes, etc.

RECOMMENDATION SYSTEMS: HOW DOES IT WORK?

Based on this, we can distinguish between three types of recommendation systems:

- CONTENT BASED SYSTEMS

Based on characteristic information

- COLLABORATIVE FILTERING SYSTEMS

Based on item-user interactions

- KNOWLEDGE BASED SYSTEMS

Based on explicitly defined set of recommendation rules

- HYBRID SYSTEMS

Combine the two types of information in order to avoid problems generated by working with only one type

TYPES OF RECOMMENDATION SYSTEMS

RECOMMENDATION SYSTEMS

- Content Based Methods
- Collaborative Filtering Methods
- Knowledge Based Methods
- Hybrid Methods

CONTENT BASED METHODS: HOW IT WORKS?

Tries to guess the features or behavior of a user given the item’s features, he/she reacts positively to.

Requirements

- some information about the available items such as the genre ("content")
- some sort of user profile describing what the user likes (the preferences)

Tasks:

- learn user preferences
- locate/recommend items that are "similar" to the user preferences "show me more of the same what I've liked"

Most Content Based recommendation techniques were applied to recommending text documents.

Advantages

01 Model don't need data about other users.

02 It is easier to scale to a large number of users.

03 The model can capture the specific interests of a user

Disadvantages

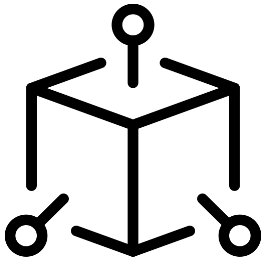
01 This technique requires a lot of domain knowledge.

02 The model can only be as good as the hand-engineered features.

03 The model can only make recommendations based on existing interests of the user.

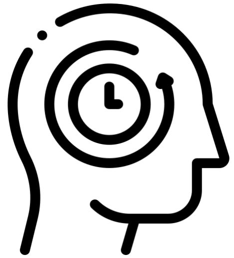
COLLABORATIVE FILTERING METHODS

Collaborative Filtering Methods



Model Based

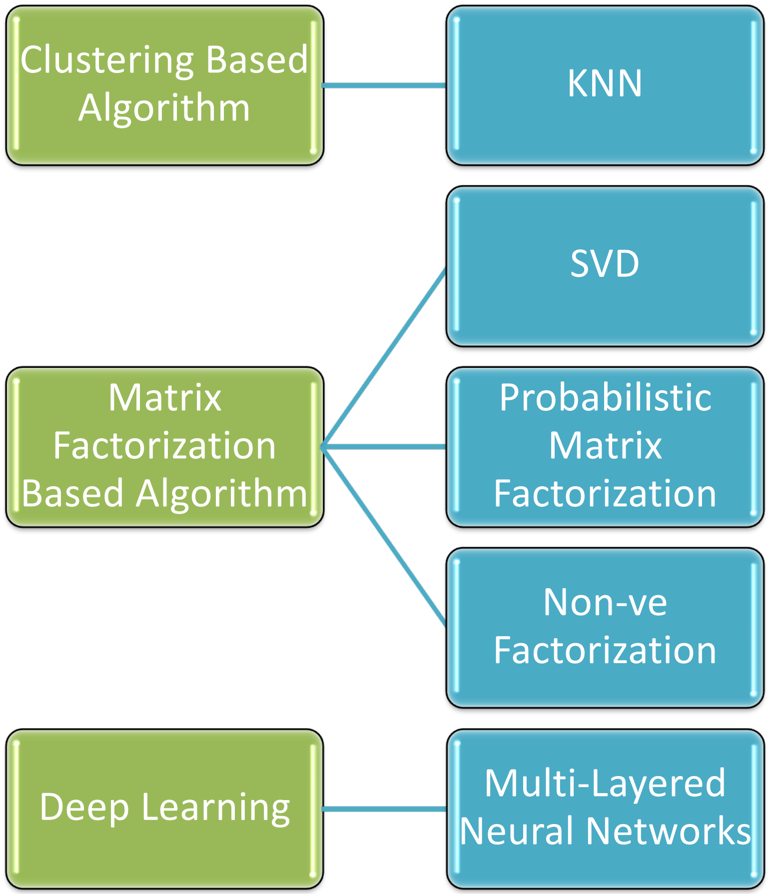
Define a model for user-item interactions where users and/or items representations must be learned from interactions matrix.



Memory Based

Define no model for user-item interactions and rely on similarities between users and/or items in terms of observed interactions.

Model Based


Clustering Based Algorithm
KNN
SVD
Matrix Factorization Based Algorithm
Probabilistic Matrix Factorization
Non-ve Factorization
Deep Learning
Multi-Layered Neural Networks

DRIVES
Development and Research on Innovative Vocational Education Skills

Model Based

Advantages

Dimensionality reduction, deals with missing/sparse data

Sparse data - percentage of the variable's cells **do** not contain actual **data**. Such "empty," or NA, values take up storage space in the file.

Disadvantages

Inference is untracable because of hidden or latent factors

A latent variable is a variable that cannot be observed. The presence of latent variables, however, can be detected by their effects on variables that are observable

DRIVES
Development and Research on Innovative Vocational Education Skills

Memory Based

User-Item

Takes a particular user, find users that are similar to that user based on similarity of ratings, and recommend items that those similar users liked

"Users who liked this item also liked …"

VS

Item-Item

Takes an item, find users who liked that item, and find other items that those users or similar users also liked

"Users who are similar to you also liked …"

Memory Based

Advantages

Easy creation and explainability of results

Disadvantages

Reduce Performance when data is sparse

Not scalable

Application Domains

Expensive Items, Not frequently purchased (car, house)

Time spans important (technology)

Explicit requirements of user (vacation)

Collaborative filtering unusable – not enough data

KNOWLEDGE BASED METHODS: LIMITATIONS

Cost of knowledge acquisition

Accuracy of models

Collaborative filtering models preference implicitly

Independence assumption can be challenged

Preferences are not always independent from each other

PROS AND CONS

	PROS	CONS
Collaborative	No knowledge-engineering effort, serendipity of results, learns market segments	Requires some form of rating feedback, cold start for new users and new items
Content-based	No community required, comparison between items possible	Content descriptions necessary, cold start for new users, no surprises
Knowledge-based	Deterministic recommendations, assured quality, no cold-start, can resemble sales dialogue	Knowledge engineering effort to bootstrap, basically static, does not react to short-term trends

HYBRID METHODS

Combines different recommendation techniques in order to obtain a better optimization of the system to avoid some limitations and problems of pure recommendation systems

The combination of approaches can be done in any of the following ways:

01 separate implementation of algorithms and combination of the result.

02 using some collaborative filtering in a content-based approach.

03 using some content-based filtering in a collaborative approach.

04 creating a unified recommendation system that brings together both approaches.

Process:

1. Combines the results of different recommenders to generate a list of recommendations or forecast
2. It is based on content and collaborative recommendation systems
3. In the beginning, both have the same weight (influence on the system)
4. Over time the weights are adjusted, according to the forecasts made

Main Benefits: All the strengths of each of the systems are used in the recommendation process

Process

1. Switch to one of the recommendation techniques according to a heuristic that reflects the ability of the recommender to produce a good rating.
2. It has the ability to avoid method-specific problems by switching to a collaborative recommendation system.

✓ **The system is sensitive to the strengths and weaknesses of its constituent recommenders.**

✗ **Usually introduces more complexity into the recommendation process, because the criteria for change, which usually increases the number of parameters in the recommendation system, has to be determined.**

Process

1. It applies an iterative refinement process to build an order of preference between different items.

2. The recommendations of one technique are refined by another recommendation technique.

3. The first recommendation technique produces a rough list of recommendations which, in turn, is refined by the next recommendation technique.

The hybridization technique is very efficient and noise-tolerant due to the coarse to fine nature of iteration.

DRIVES
Development and Research on Innovative Vocational Education Skills

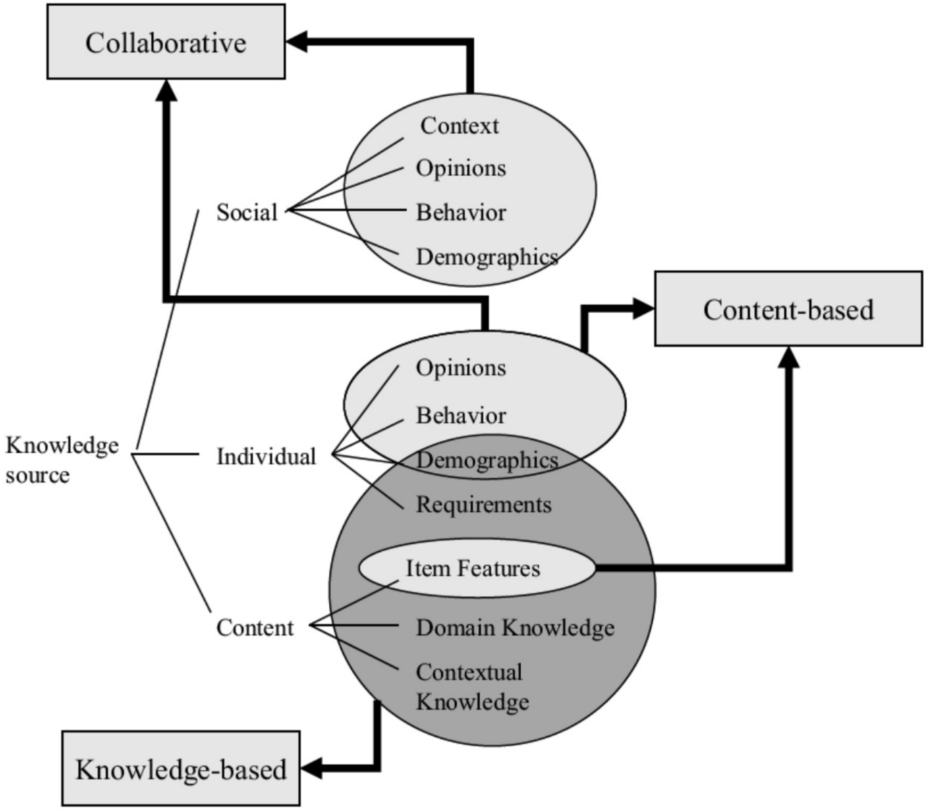
Process

1. They combine recommendation results from different recommendation techniques at the same time, instead of having only one recommendation per item.

2. Each item has multiple recommendations associated with it from different recommendation techniques.

3. In mixed hybridization, individual performances do not always affect the overall performance of a local region.

KNOWLEDGE SOURCES AND RECOMMENDATION TYPES

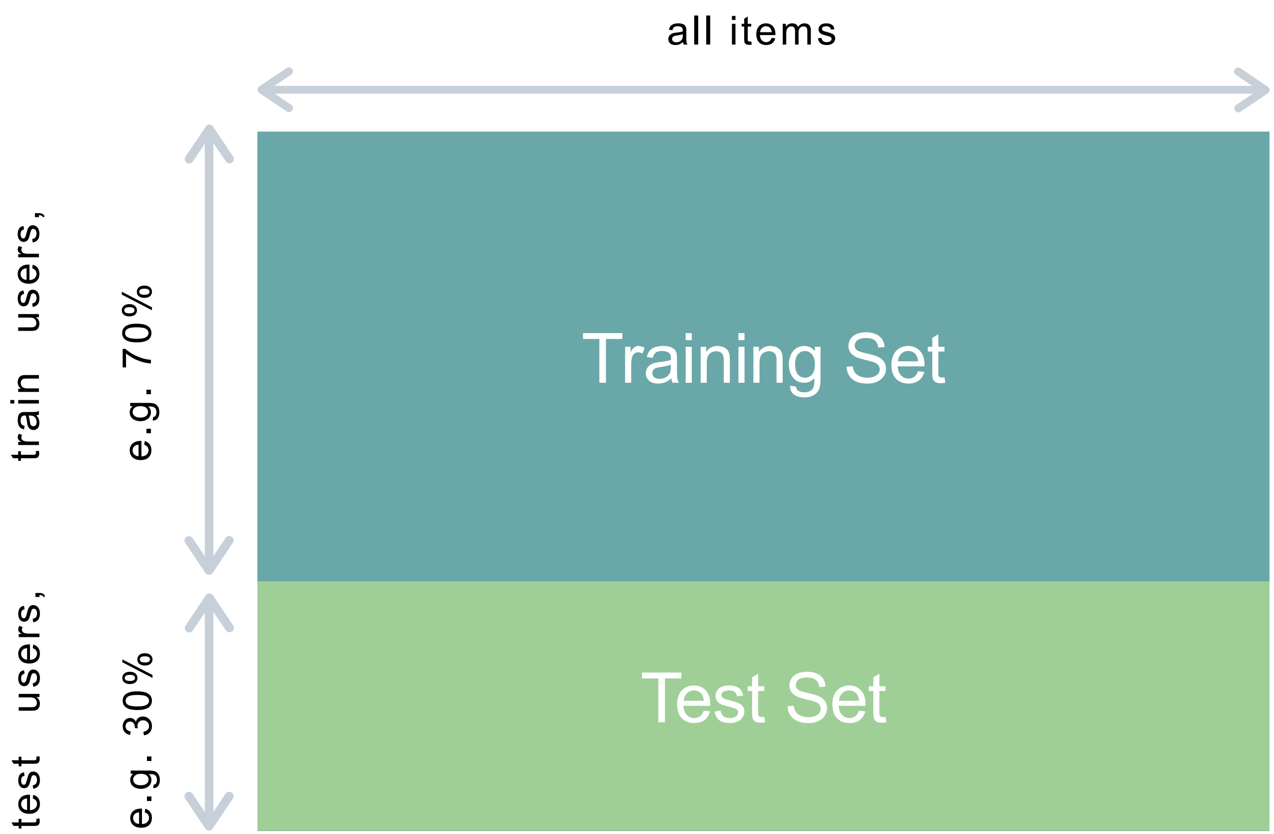


ROLE OF MACHINE LEARNING IN RECOMENDATION SYSTEMS

Recommendation Systems are composed by Machine Learning algorithms that generate recommendations of a given type of item for users. Recommender systems can be evaluated similarly as classical Machine Learning models.

Interactions of randomly selected testing users are cross validated to estimate the performance of recommender on unseen ratings.

To validate the recommendation system, one may divide users into:



EXERCISE

<describe an exercise for students by which they can prove the skills>

- There are 4 types of reccommendation methods: Content Based Method, Collaborative Filtering Methods, Knowledge Based Methods and Hybrid Methods;
- All the methods have advantages and disadvantages and are very useful for different use cases;
- Recommendation systems work with two types of information: Characteristic Information and Item-User Interactions;
- Recommendation Systems are composed by Machine Learning algorithms that generate recommendations of a given type of item for users.

REFERENCES

- Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325-341). Springer, Berlin, Heidelberg.
- Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, *16*(3), 261-273.
- Khanal, S. S., Prasad, P. W. C., Alsadoon, A., & Maag, A. (2019). A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies*, 1-30.
- Naumov, M., Mudigere, D., Shi, H. J. M., Huang, J., Sundaraman, N., Park, J., ... & Smelyanskiy, M. (2019). Deep learning recommendation model for personalization and recommendation systems. *arXiv preprint arXiv:1906.00091*.
- Levy, K. L., & Lofgren, N. E. (2010). *U.S. Patent Application No. 12/764,091*.
- Son, J., & Kim, S. B. (2017). Content-based filtering for recommendation systems using multiattribute networks. *Expert Systems with Applications*, *89*, 404-412.
- Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender systems handbook* (pp. 73-105). Springer, Boston, MA.

REFERENCES

- Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). *Collaborative filtering recommender systems*. Now Publishers Inc.
- Thorat, P. B., Goudar, R. M., & Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, *110*(4), 31-36.
- Chen, A. Y. A., & McLeod, D. (2006). Collaborative filtering for information recommendation systems. In *Encyclopedia of E-Commerce, E-Government, and Mobile Commerce* (pp. 118-123). IGI Global.
- Bouraga, S., Jureta, I., Faulkner, S., & Herssens, C. (2014). Knowledge-based recommendation systems: a survey. *International Journal of Intelligent Information Technologies (IJIIT)*, *10*(2), 1-19.
- Lin, C. Y., Wang, L. C., & Tsai, K. H. (2018). Hybrid real-time matrix factorization for implicit feedback recommendation systems. *IEEE Access*, *6*, 21369-21380.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, *12*(4), 331-370.

REFERENCE TO AUTHORS

Diana Ferreira

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

iD [0000-0003-2326-2153]

Regina Sousa

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

iD [0000-0002-2988-196X]

José Machado

- Associate Professor with Habilitation at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

iD [0000-0003-4121-6169]

REFERENCE TO AUTHORS

António Abelha

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

iD 0000-0001-6457-0756

Victor Alves

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

iD 0000-0003-1819-7051

REFERENCE TO AUTHORS

This Training Material has been certified according to the rules of **ECQA – European Certification and Qualification Association.**

The Training Material was developed within the international job role committee **“Machine Learning Engineer”:**

UMINHO – University of Minho (https://www.uminho.pt/PT)

The development of the training material was partly funded by the EU under Blueprint Project DRIVES.

DRIVES

Development and Research on Innovative Vocational Education Skills

Thank you for your attention

DRIVES project is project under **The Blueprint for Sectoral Cooperation on Skills in Automotive Sector**, as part of New Skills Agenda.

The aim of the Blueprint is **to support an overall sectoral strategy and to develop concrete actions to address short and medium term skills needs.**

Follow DRIVES project at:

More information at:

www.project-drives.eu

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.