Allview

D1.4 – ALLVIEW AI/ML algorithms

Artificial Intelligence to connect people, jobs and training

Version – March 2022

www.allview.eu

621192-EPP-1-2020-1-ES-EPPKA3-VET-COVE

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

This project has been funded with the support of the
Erasmus+ programme of the European Union

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

Content

Allview

1

Introduction

1. Introduction

This document summarizes the results obtained from task T1.4 along the period M8-M18. During that period, the partner leader of T1.4 and D1.4, UPCT, has designed and developed the platform engine based on Artificial Intelligence (AI) and Machine Learning (ML) algorithms. This technology has offered a front-end with a personal Learning and Development (L&D) portal to the ALLVIEW platform thanks to use of Recommender Systems (RS), a branch of AI/ML. The developed AI/ML algorithms have taken user data as inputs of the system to report some user recommendations as outputs of the platform.

As it was defined in the Deliverable D1.3, this RS uses some user data like the user preferences, the user skills and competences, and similar user profiles to make recommendations to users about suitable learning paths, training courses or job positions. The platform also shows the reputation of training courses performed by registered users on the platform and companies that offer job positions based on people opinions. This information may be valuable information for potential recruiters. Other features of the ALLVIEW platform are the automatic (mostly ML powered) collect of the learning interests of the people and build a database where from macroscopic statistics could be analyzed, like people who know for X typically is interested in Y, or Z skill is getting a lot of interest in society or within a group. In the same way an enterprise (user Company) may know if there are enough people recruitable in a particular skill, or if it will have difficulties to find people.

AI and ML are two concepts related between them, but different (see Figure *1*). AI is a technology which enables a machine to simulate human behavior with the goal of making a smart computer system like humans to solve problems, and ML is a kind of AI technique which allows the system to obtain knowledge with no explicit programming. In other words, ML is a subset of AI which allows a machine to automatically learn from past data without programming explicitly. The third concept is Deep Learning (DL), a technique for implementing ML algorithms. RS are a type of ML technique used to help users find new items or services. RS are the focus of this report since it is used in the core of the ALLVIEW platform.

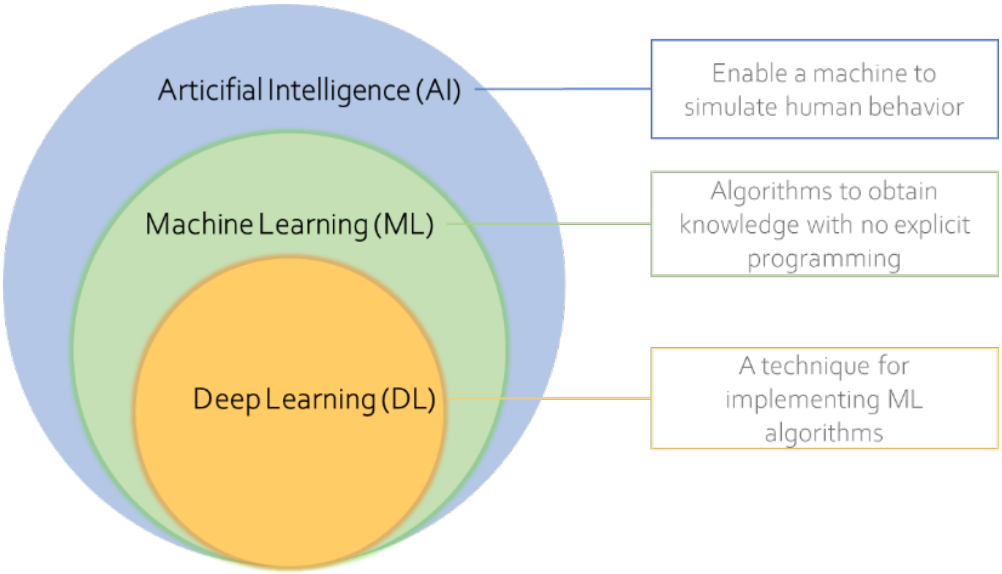


Figure 1. The relationship between AI, ML and DL [1]

Below, in this report, a state of the art of the AI and ML algorithms is presented. Next sections are focused on RS because, as it is detailed in the document, these are the algorithms used in this type of applications. About RS, two techniques are identified to be applied on ALLVIEW platform, content-based and collaborative filtering techniques. The developed algorithms using the mentioned techniques are also presented. After that, the chapter 4 presents the evaluation of ML techniques with data sets. For each defined use case, a training set and the expected output have been detailed showing also the real results obtained. In chapter 5, the final deployment is presented including the cloud solution provider and the structure of the public repository, where the open-source code has been allocated. Finally, the conclusions are summarized.

1 Source: https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications

Allview

2

State of the art of AI/ML techniques

2. State of the Art of AI/ML Algorithms

AI and ML is a field in continuous growing with new applications that make easy human tasks or decisions. There are different techniques depending on the issue to be solved. Some techniques require to use the input data, train the system and get the output data, and others learn by themselves using data from user experience. Details of how each technique or algorithm works is important since not all AI/ML algorithms are suitable for solving all issues.
This section describes a state of the art of AI/ML algorithms. Figure 2 shows graphically how ML algorithms are classified.

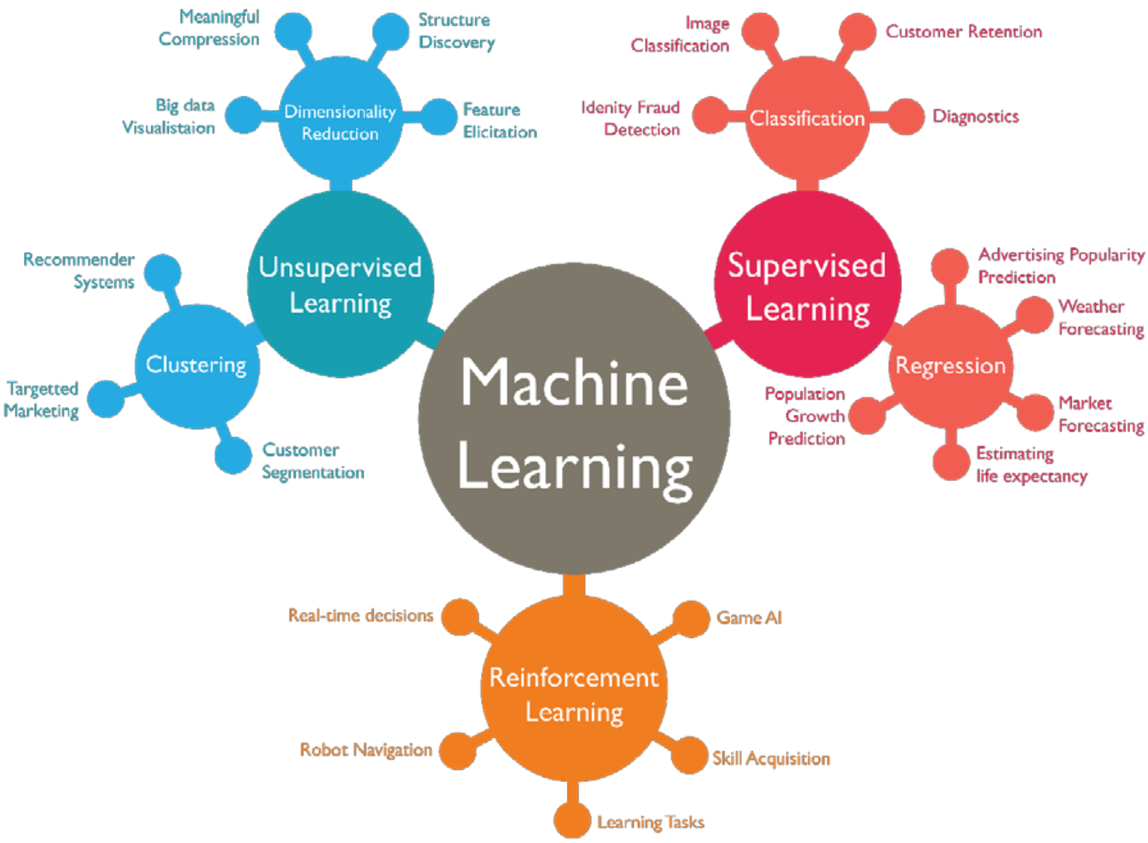


Figure 2. ML techniques classification 2

ML algorithms can be mainly classified on three big categories:

- Supervised ML algorithms use learned knowledge from previous and present data with the help of labels to predict events. The labelled data means some input data is already tagged with the correct output. This requires an adequate training process of dataset to foresee the output values. After this process, the system can provide results to an input data. The algorithm compares the obtained results with the actual and expected results to identify errors to change the model adjusting the error based on results.
- Unsupervised ML algorithms are used when the training data is non-classified and not labelled and not require any training process. These algorithms analyse how the

2 Source: https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications

system can obtain a function to explain patterns from the unlabeled data. The system does not identify the proper output, but it discovers the data and writes observations from dataset to find hidden patterns from unlabeled data. RS is classified as unsupervised algorithm and it can be defined as a learning technique where users can customize their sites to meet customer's preferences. For example, online users can get a rating of a product or service and related items when they search an item thanks to the existing RS. RS allows users to find products or information in a different way.

- Reinforcement Learning (RL) algorithms enable an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences. In this case, the algorithm continuously learns from its environment using iteration.

Table *1* shows a summary of the most popular ML algorithms. There is such quantity of algorithms that it is not possible to detail all in this study. Next sections describe those that are of interest for this work.

Table 1. ML algorithms

MACHINE LEARNING ALGORITHMS

Supervised algorithms	Unsupervised algorithms	Reinforcement learning
Linear regression*	Clustering*	Q-learning
Multiple regression*	K-means clustering	SARSA (State-Action-Reward-State-Action)
Logistic regression*	KNN (k-nearest neighbors) *	Deep Q-Networks
Polynomial regression	Recommender systems*	DDPG (Deep Deterministic Policy Gradient)
Non-linear regression	Anomaly detection	
Classification	Neural networks	
KNN (k-nearest neighbors)	Dimension reduction	
Naive Bayes theorem	Density estimation	
Linear discriminant analysis	Principle component analysis	
Learning vector quantization	Independent component analysis	
Support vector machine*	A priori algorithm	
Decision trees*	Singular value decomposition	
Random forest	Market basket analysis	
Boosting		
AdaBoost		

* It is detailed in next sections

2.1 Supervised Learning Algorithms

Taking as reference the schema showed in Figure 3, supervised learning models learn about each type of data because they are trained using labelled dataset. In the example of this picture, the first step is to train the model for each shape (star, circle and square). After training process, the model is tested using a test set (a subset of the training set) to verify that the identification of the shapes is correct, and finally, the model is ready to make output predictions.

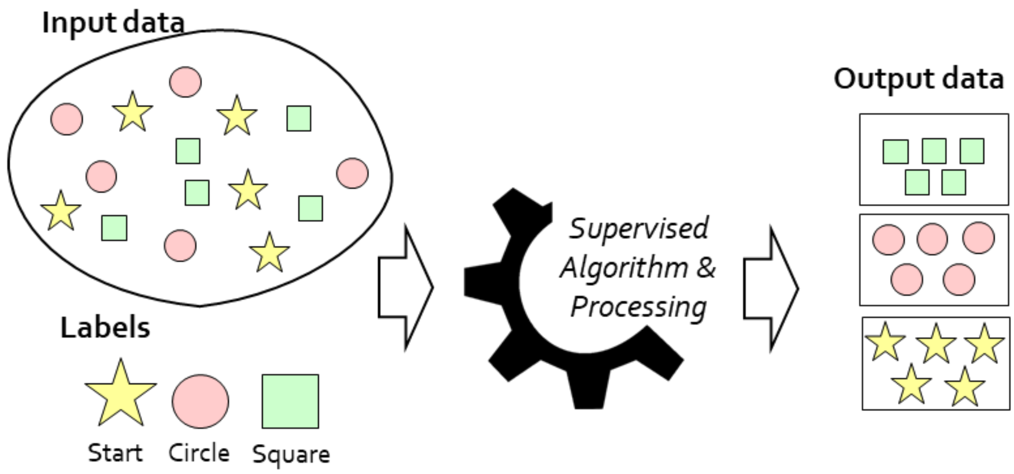


Figure 3. Supervised Learning

Supervised learning algorithms can be classified into two types: regression and classification.

- Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as weather forecasting, market trends, etc.
- Classification algorithms are used when the output variable is categorical, which means there are two classes such as yes/no, male/female, true/false, etc.

Supervised learning algorithms can help to solve real issues like spam filtering, but it is not valid for managing complex tasks. These algorithms are useful when there is a training dataset but cannot predict the correct output if the test data is different from the training dataset. In general, enough knowledge about the classes of object is required, because if the dataset is incorrect, the algorithm could learn incorrectly which can bring losses.

Linear regression

Linear regression algorithm assumes a linear relationship between an independent variable, Input (X), and a dependent variable, Output (Y) (see Figure 4). When the algorithm receives data, it uses the function, calculates and maps the input to a continuous value for the output. A linear regression line has an equation of the form Y = a + bX. This analysis is used to predict the value of a variable based on the value of another variable. The variable to be predicted is called the dependent variable.

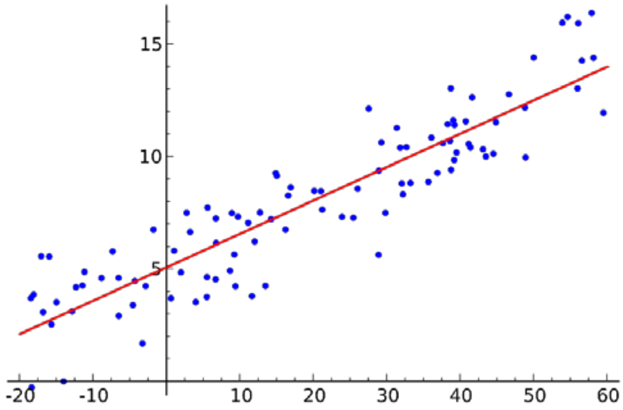


Figure 4. Linear Regression. X axis: independent variables. Y axis: dependent variables 3

Multiple regression

Multiple regression is like linear regression, but with more than one independent value, meaning that we try to predict a value based on two or more variables.

Logistic regression

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. This algorithm predicts discrete values for the set of independent variables that have been passed to it. As Figure 5 shows, the target or dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

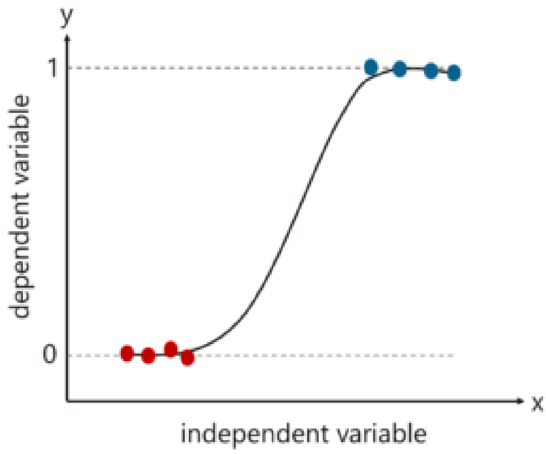


Figure 5. Logistic Regression 4

3 Source: https://upload.wikimedia.org/wikipedia/commons/3/3a/Linear_regression.svg

4 Source: https://www.edureka.co/

Support Vector Machine

The main goal of Support Vector Machine (SVM) technique is to divide the datasets into classes to find a maximum marginal hyperplane (MMH). This model represents the classes in a hyperplane in multidimensional space (see Figure 6). The hyperplane is generated in an iterative manner by SVM so that the error can be minimized. There are 3 important concepts in this technique:

- Support Vectors are data points that are closest to the hyperplane. Separating line is defined with the help of these data points.
- The Hyperplane is a space which is divided between a set of objects having different classes.
- The Margin is defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin.

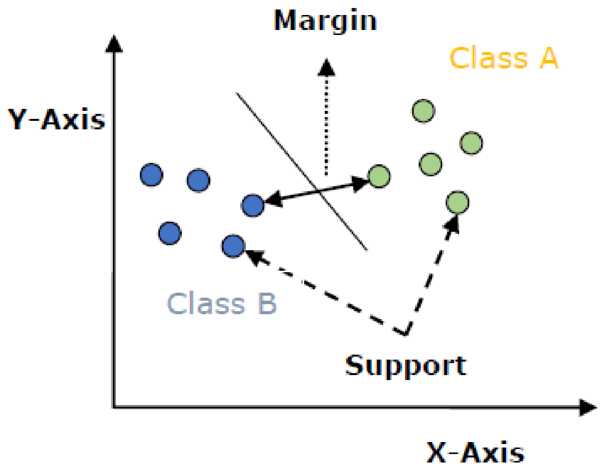


Figure 6. Support Vector Machine 5

Decision Tree

Decision Tree is a type of Supervised ML technique where the data is continuously split according to a certain parameter, and it can be used for both classification and Regression problems. It is a graphical representation for getting all possible solutions to a problem/decision based on given conditions. A decision tree asks a question and based on the answer (Yes/No), it split the tree into subtrees (see Figure 7). The tree is divided into 3 entities: internal nodes represent the features of a dataset; branches represent the decision rules and each leaf node represents the outcome.
There are two main types of Decision Trees:

- Classification trees (Yes/No types).
- Regression trees (Continuous data types)

5 Source: https://www.tutorialspoint.com/

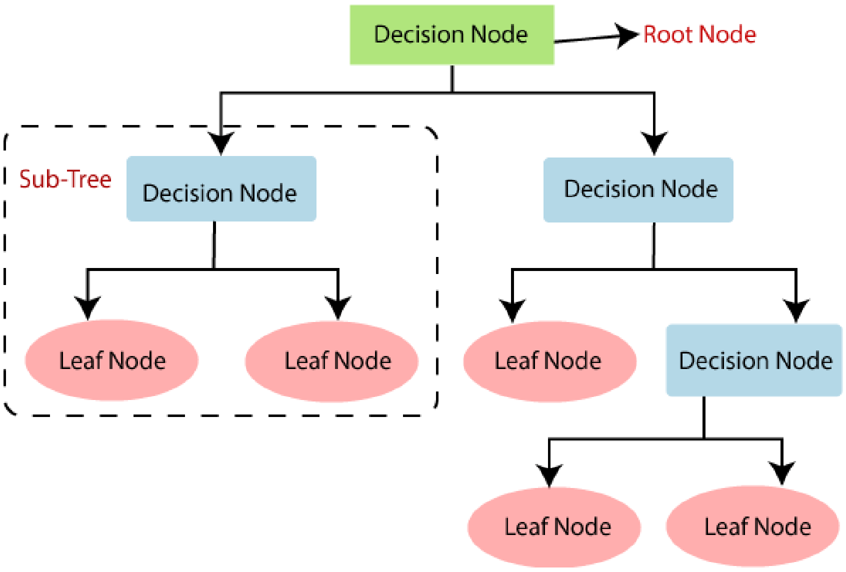


Figure 7. Decision tree 6

This technique has some advantages and disadvantages. The most known advantage is that it is a simple method where the process is like a human reasoning when we make any decision in real life. It can be very useful for solving decision-related problems and it also helps to think about all the possible outcomes for a specific problem. On the other hand, there is a disadvantage that must be considered. This technique could be very complex in case of the decision tree contains lots of layers.

2.2 Unsupervised Learning Algorithms

In the previous section it mentions that models in supervised learning algorithms are trained using labelled data under the supervision of training data, but it is not possible on all sceneries. On some occasions, there is no labelled data and the process requires to find the hidden patterns from the given dataset, and here is where unsupervised learning algorithms are applied (see Figure 8).

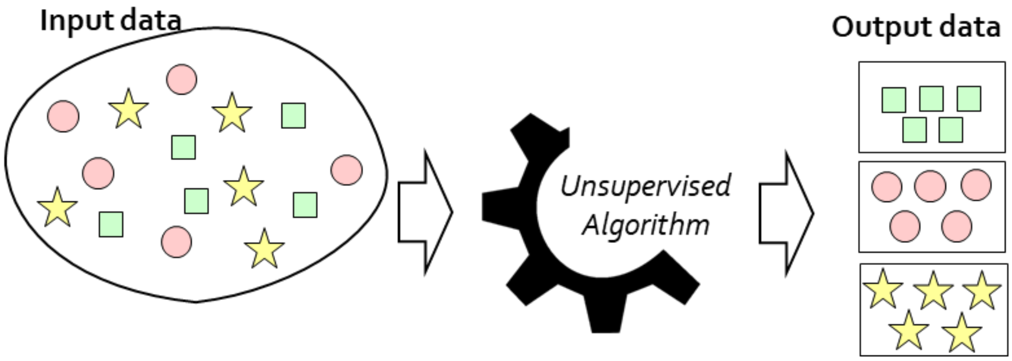


Figure 8. Unsupervised Learning

6 Source: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

Unsupervised learning algorithms can be further categorized into two types: clustering and association.

- Clustering is a method of grouping the objects into clusters. The main idea is to create a group with the objects with most similarities and another group with the objects that have less or no similarities with the first. Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those points in common.
- Association is a rule used for finding the relationships between variables in a large database. It determines the set of items that appear together in the dataset. As example, this type of algorithms makes marketing strategy more effective, because people who buy X item could also purchase Y item. A typical example of Association rule is Market Basket Analysis.

Unsupervised learning is used for more complex tasks as compared to supervised learning, but it is intrinsically more difficult than supervised learning as it does not have corresponding output. Another disadvantage is that the result of the unsupervised learning algorithm might be less accurate as input data is not labelled, and algorithms do not know the exact output in advance.

Clustering

Clustering is an unsupervised machine learning task that interpret the input data and find natural groups or clusters in feature space. It is basically a collection of objects based on similarity and dissimilarity between them. For instance, we can identify that there are 3 clusters in Figure 9. There are not specific rules to classify when a clustering algorithm is good, since this depends on the user criteria. It means, users should find the right criteria they may use which satisfy their need.

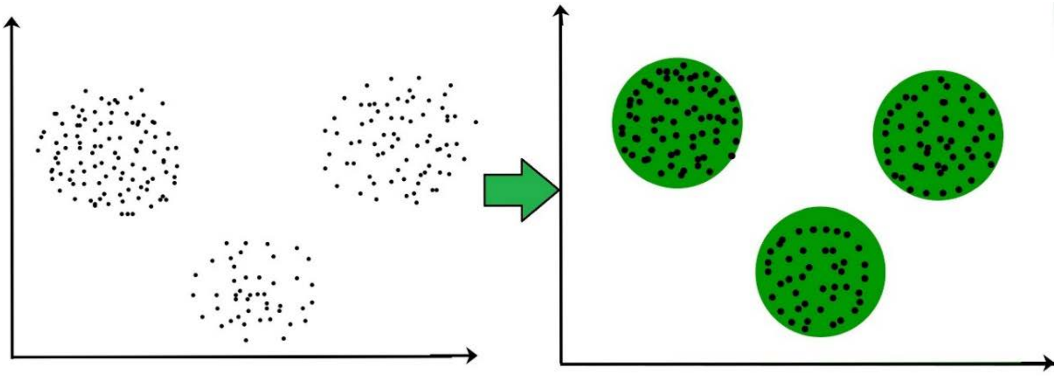


Figure 9. Clustering 7

There are many clustering methods, but K-means clustering algorithm is the simplest unsupervised learning algorithm that solves clustering problem. K-means algorithm partitions n observations into k clusters where each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster.

7 Source: https://www.geeksforgeeks.org/

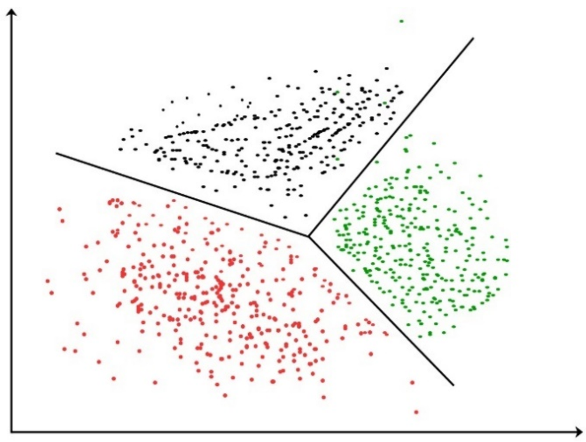


Figure 10. K-means clustering algorithm

K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) algorithm is a simple and easy-to-implement supervised ML technique, that can be used to solve both classification and regression problems. This algorithm assumes that similar things are near to each other. It means, the algorithm assumes the similarity between new data and available data and put the new data into the category that is most like the available categories. K-NN algorithm is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. At the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is more similar to the new data.

Considering the example in the Figure *11*, there are 2 categories (Category A and Category B), and there is a new data point x_n, so this data point will lie in which of these categories. The objective of K-NN is to identify the category or class of a particular dataset.

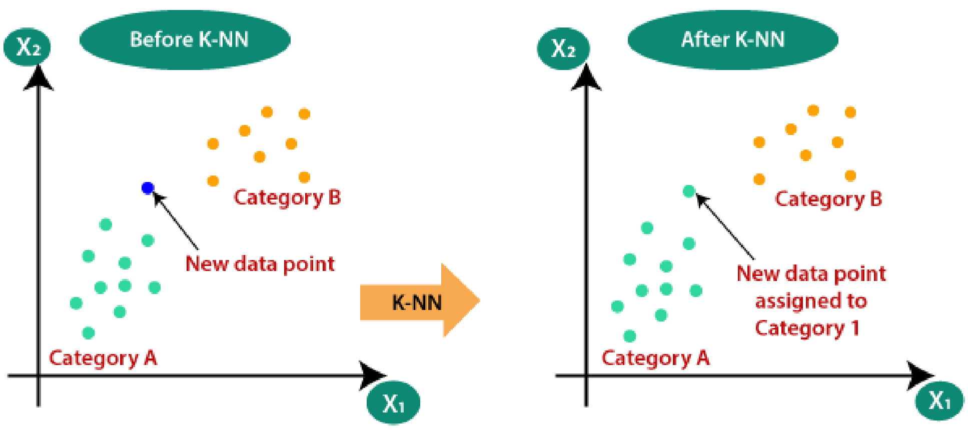


Figure 11. K- Nearest Neighbour algorithm *8*

8 Source: https://www.javatpoint.com/

K-NN algorithm is simple to implement and robust to the noisy[9] training data, although it can be more effective if the training data is large. However, this technique may be complex in some cases calculating the value of K, even the computation cost can be high due to calculate the distance between the data points for all the training samples.

Recommender Systems

There is a huge amount of information on the Internet, so it is not always easy to access the content that users are interested in. RS are information filtering systems that manage data to create fragments according to user's preferences, interests, or observed behavior about a particular item or items. A RS has the ability to predict whether a particular user would prefer an item or not based on the user's profile and its previous information. One of the aims of the recommendations is to speed up the searches and make it easier for users to access the content they have always been interested in, and surprise them with several offers they would have never searched for.

RS algorithms allow the system to make recommendations to users based on user's preferences, ratings or other user's preferences with similar profiles. There are mainly two approaches (see Figure 12): content-based recommendation and collaborative recommendation.

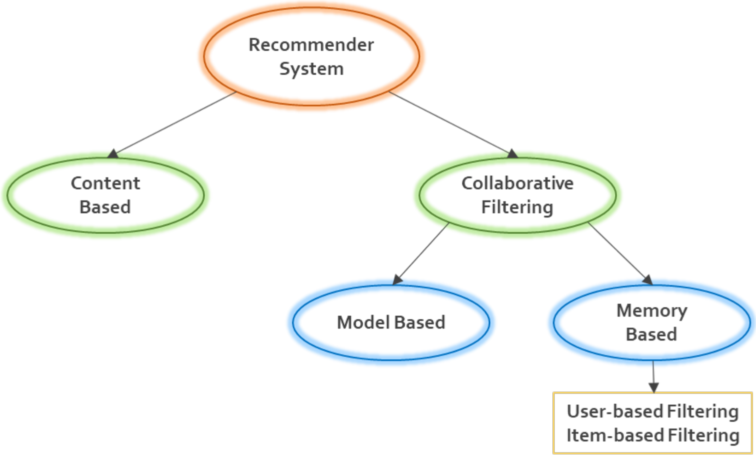


Figure 12. Types of Recommender Systems

Content-Based Filtering

Content based (CB) recommender methods are used to decide the outcomes based on product similarities instead of user feedback or interaction. A description of the item and a profile of the user's preferences are required to make recommendations accordingly (see Figure 13). These methods are best suited to situations where there is known data on an item (training course, description, etc.), but not on the user. In this method, keywords are used to describe the items and a user profile is built to indicate the type of item this user likes. It means that these algorithms try

9 In AI, the word "noise" refers to the quality of training data. k-NN classifier is usually featured as noise-sensitive. ts accuracy highly depends on the quality of the training data.

to recommend items that are similar to those that a user liked in the past or is examining in the present.

There are some advantages and disadvantages to be considered. The advantages are: 1) the model does not need any data about other users, since the recommendations are specific to this user. This makes it easier to scale to many users, 2) the model can capture the specific interests of a user, and can recommend items that very few other users are interested in. On the other hand, the model can only make recommendations based on existing interests of the user. In other words, the model has limited ability to expand on the users' existing interests.

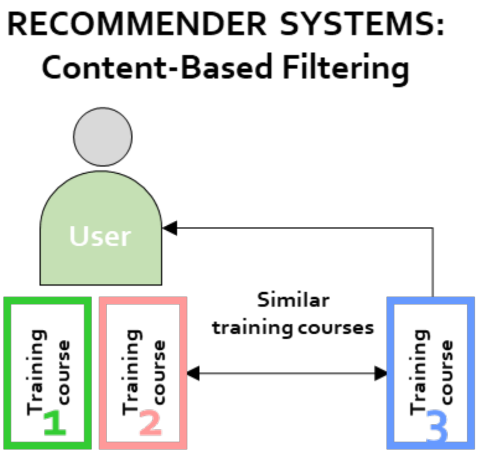


Figure 13. Recommender Systems – Content-Based filtering

Collaborative Filtering

Collaborative Filtering (CF) RS are based on user past behavior as well as user similar decisions made by other users. In CF method only user content and profile information are not enough. The rating's users are associated with other user's behavior giving a similar rating. It finds similarities between users and items to make assumptions for missing rating values and deducing new recommendations. The model can help users to discover new interests. In isolation, the ML system may not know the user is interested in a given item, but the model might still recommend it because similar users are interested in that item.

In Figure 14, the system finds two users with similar profiles (same interests). The user 1 is interested on training courses 1, 2 and 3. The user 2 is interested on training courses 1 and 2. So, the RS detects that user 1 and user 2 are similar, and RS suggests the training course 3 to user 2.

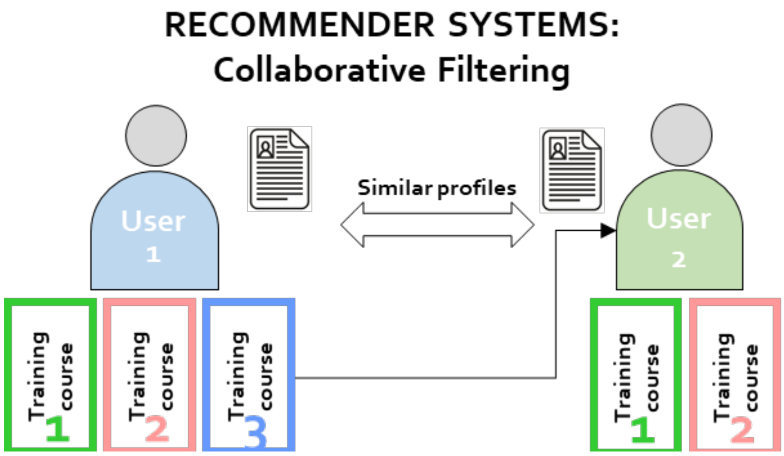


Figure 14. Recommender Systems – Collaborative filtering

Beyond this, there are two types of CF methods: memory-based and model-based collaborative filtering methods.

Memory-based Collaborative Filtering

Memory-based CF calculates the similarity between users or items using the user's previous data based on ranking. The main objective of this method is to describe the degree of resemblance between users or objects and discover homogenous ratings to make suggestions.
There are two types of Memory-based CF two methods (see examples in Figure *15*):

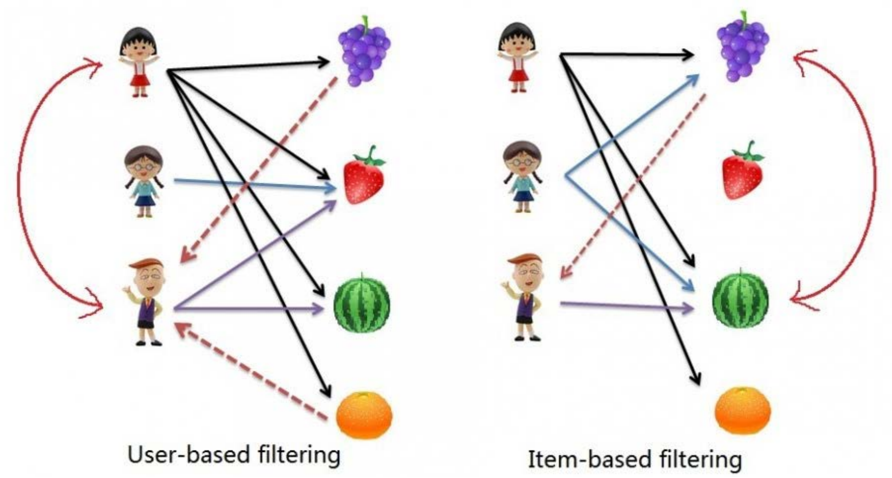


Figure 15. Memory-based Collaborative filtering methods *10*

a) User-based Collaborative Filtering (UB-CF): the algorithm predicts the items that a user might like based on ratings given to that item by the other users who have similar taste with that of the target user
b) Item-based Collaborative Filtering (IB-CF): the algorithm finds the same items that the target user has already viewed.

10 Source: https://medium.com/

Model-based Collaborative Filtering

Model-based Collaborative Filtering models are used to forecast and calculate how a user gives a rating to each item. These algorithms are based on ML to predict unrated products by user ratings. These algorithms are further divided into different subsets, i.e., Matrix factorization-based algorithms, deep learning methods, and clustering algorithms.

2.3 Reinforcement Learning Algorithms

As compared to unsupervised learning, RL has different goals. While the goal in unsupervised learning is to find similarities and differences between data points, in reinforcement learning the goal is to find a suitable action model that would maximize the total cumulative reward of the agent. In this algorithm, for every result obtained the algorithm gives feedback to the model under training. Additionally, the agent learns automatically using feedbacks without any labelled data, unlike supervised learning, therefore the agent is bound to learn by its experience only.

The Figure *16* represents the elements involved in the performance of the RL model: 1) Environment is the physical world in which the agent operates, 2) State is the current situation of the agent, and 3) Reward is the feedback from the environment. The agent interacts with the environment and explores it by itself. The primary goal of an agent in RL is to improve the performance by getting the maximum positive rewards.

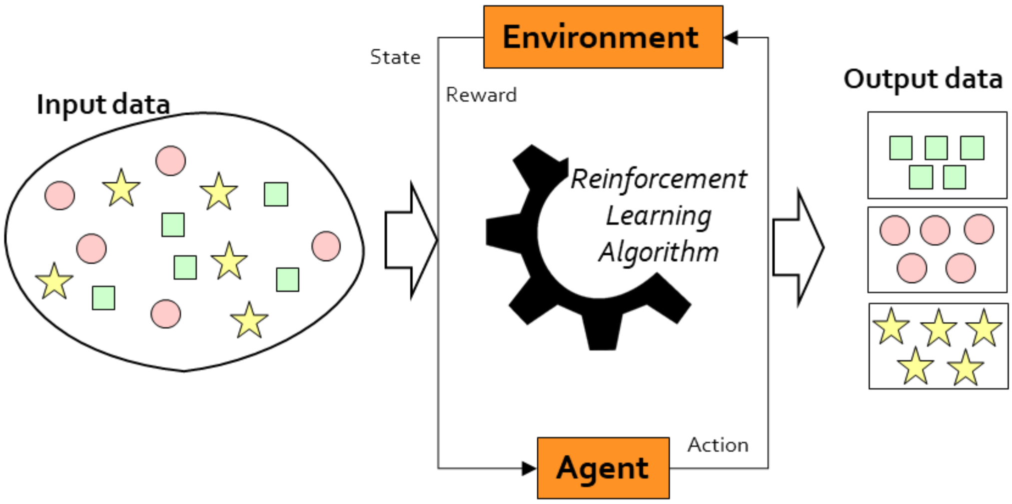


Figure 16. Unsupervised Learning

This technique is very similar to the human learning technique, and thanks to the training process almost any errors could be solved. However, one disadvantage of this learning is that the model requires a lot of training data to develop accurate results, it consumes time and lots of computational power.

2.4 Keyword Extraction technique

Keyword Extraction is another technique used on ALLVIEW platform. This technique is the automated process of extracting the most relevant words and expressions from text. It is a text analysis technique that automatically extracts the most used and most important words and expressions. It helps summarize the content of texts and recognize the main topics discussed. And, not only that, but this technique also uses ML/AI with Natural Language Processing (NLP) to break down human language so that it can be understood and analyzed by machines.

There are different approaches to keyword extraction, and this section is focused on one of the ML-based models. Statistical approaches are one of the simplest methods for identifying the main keywords within a text. There are several types of statistical approaches: word frequency, word collocations and co-occurrences, TF-IDF (short for term frequency–inverse document frequency), and RAKE (Rapid Automatic Keyword Extraction). There is no technique better than other, it is depending on the aim of the task. For this platform, the technique used is TF-IDF, leaving the others out of the scope of this document.

TF-IDF (short for term frequency–inverse document frequency), is a statistical measure that evaluates how relevant a word is to a document in a collection of documents. This is done by multiplying two metrics: how many times a word appears in a document, and the inverse document frequency of the word across a set of documents [11]. One of the most important uses is in automated text analysis and is very useful for scoring words in ML algorithms for NLP. In fact, ALLVIEW platform uses Keyword Extraction to find keywords on user profiles (from ESCO occupations descriptions) or job positions offers to help make better recommendations to users.

[11] Source: https://www.kaggle.com/rowhitswami/keywords-extraction-using-tf-idf-method

3

Identification of ML techniques to be applied

3. Identification of ML techniques to be applied

In order to identify the ML technique to be applied in ALLVIEW platform, knowing how to run the platform is key to select the most suitable technique. Firstly, a brief description about the main functionalities planned with ML to offer recommendations to users is introduced. Secondly, considering the ALLVIEW platform engine and the ML techniques detailed in the previous chapter, a decision about the ML algorithm selected will be made.

As we already know, ML help to provide functionalities like, e.g.: personalized learning programs for target people. That is, a front-end with a personal Learning and Development (L&D) portal, using AI-powered RS for learning resources built on top of data.
As project description states, the input to the platform is information about employee profile, target interests or goals (e.g., re-skill, deepen existing skills). The output is the recommend L&D actions (courses, etc.). All these features built on top of AI-powered data from past experiences of others (with more weight if profile & goal are similar). Moreover, the tool use feedbacks of all type of users to AI-powered reputation system for learning resources. Reputation is built automatically from, i) opinions on people enrolled/performing training, ii) rates of success in the short, medium and long term of people performing training, iii) grades and others (to define). The output helps to RS, which assists the decision of the target "learner".

Feedbacks to AI-powered evaluation of fulfilment of learning outcomes, may be valuable information for potential recruiters. The output will be automatic evaluation of potential workers, or suitability for a particular job, coming from its trajectory in L&D as appears in the system.
Other features of the open-source software platform are that the platform automatically (mostly ML powered) collect the learning interests of the people, and build from there macroscopic statistics: e.g., people who know for X typically is interested in Y, or Z skill is getting a lot of interest in society or within a group. In the same way, an enterprise may know if there are enough people recruitable in a particular skill, or else, if it will have difficulties to find people.

Being more precise, several recommendations have been defined depending on the type of user of the ALLVIEW platform. As it was mentioned in previous deliverables in WP1, where the platform was defined and detailed, one of the newest options in the ALLVIEW platform is to include various types of recommendations to users. There are 3 types of users: People, Training Provider and Company. People users register on ALLVIEW platform to look for acquiring new skills and competences as well as new job positions. They are encouraged to fill in their profile with their personal data and preferences to receive recommendations about training courses and job positions offers based on that information, among others. Training provider users publish training courses specifying the learning outcomes. These types of users can also receive recommendations about candidates for their training courses and which the most popular skills are to prepare new and novel training courses. And the third type of user, Company users include job positions offers specifying the required skills and competences among others, and these users also can look for training courses for their employees and can receive candidate recommendations for their job position offers.

The explained features above match directly with the RS detailed in the previous section, in which two types of algorithms have been described. The ALLVIEW platform engine implements the content-based filtering algorithm as well as the collaborative filtering algorithm. Next sections, considering these techniques applied, detail each technique and the information exchange the engine needs to perform to obtain as output the established recommendations. ALLVIEW platform uses 2 ML techniques to implement a RS: content-based filtering and collaborative filtering techniques.

Apart from recommendations based on ML, the ALLVIEW platform also makes recommendations based on data collect by the platform. Table 2 includes all types of recommendations made by the ALLVIEW platform, the recommendation description, the ML technique used for each type of recommendation, specifying the type of user of the ALLVIEW platform. Next sections describe the algorithm developed for each case.

Table 2. Recommendations in ALLVIEW platform

Recommendations in ALLVIEW platform

Id	Type of user	Recommendation Description	ML	ML Technique	Example
1	People	Training courses based on user job challenges, using other similar user content and their ratings.	Yes	Content-based filtering	Figure 17
2	People	Training courses based on previous courses already done by the user.	Yes	Content-based filtering	Figure 18
3	People	Training courses based on ranking made by other users with the same interests.	Yes	Collaborative filtering	Figure 19
4	People	Training courses demanded by companies, considering the skills included in job offers. This recommendation is made to People users who do not have included any job challenge or ranked any training course.	Yes	Content-based filtering	Figure 20

5	People	Job position offers based on user profile skills. Ranking is not used.	Yes	Content-based filtering	Figure 21
6	People	Learning paths based on user challenges and current user competences.	No	-	Figure 22
7	Training Provider	Candidates for attending training courses based on keywords from job challenges.	Yes	Content-based filtering	Figure 23
8	Training Provider	Demanded skills to prepare new training courses.	No	-	Figure 24
9	Company	Training courses for Companies' employees based on previous training courses done.	Yes	Content-based filtering	Figure 25
10	Company	Training courses for Companies' employees based on other users with the same interests.	Yes	Collaborative filtering	Figure 26
11	Company	Candidate for job position offers (based on skills required in job position offers).	Yes	Content-based filtering	Figure 27
12	All	Top rated training courses.	No	IMDB weighted average method	-

In summary, the system makes recommendations about training courses, job position offers and candidates considering different criteria and using different techniques. On some occasions, in which the system considers that the calculated recommendation is not good enough, the system will use another information to make appropriate recommendations.

3.1 Making recommendations to People users

ALLVIEW platform makes several recommendations to People users based on different criteria. The recommendations are training courses, job position offers and learning paths. Training courses and job position offers recommendations are made using ML techniques. On the other hand, learning paths are made considering the user challenges and current user competences. Below, each recommendation is detailed with a visual diagram including the steps followed to achieve the final aim in each case.

Figure *17* shows the process to recommend training courses to People users based on the user job challenges, using other similar user content and considering the ratings of the training courses performed by those users. Users can include the job challenges on their user profiles once the register is done, and they also can rate a training course once it is finished. Then, this information helps ALLVIEW platform to make appropriate recommendations using the content-based filtering technique.

The first step is to search users with similar job challenges. At this point, the system applies a similarity factor between 0 and 1 to each similar user based on the similarities on the job challenges. Once the ALLVIEW platform has found similar users based on that information, the second step is to compare the training courses made by the similar users and its ratings, and to find the training courses to recommend. For this, the ALLVIEW RS create a prediction with the following equation:

$$Prediction = user\ similarity\ factor * training\ course\ rating$$

Finally, the ALLVIEW RS will recommend the training courses with higher predictions. For instance, considering the example in the Figure *17*, the ALLVIEW RS has calculated that the highest prediction factor to make the recommendation to user 2 is for the training course C already performed by the user 1.

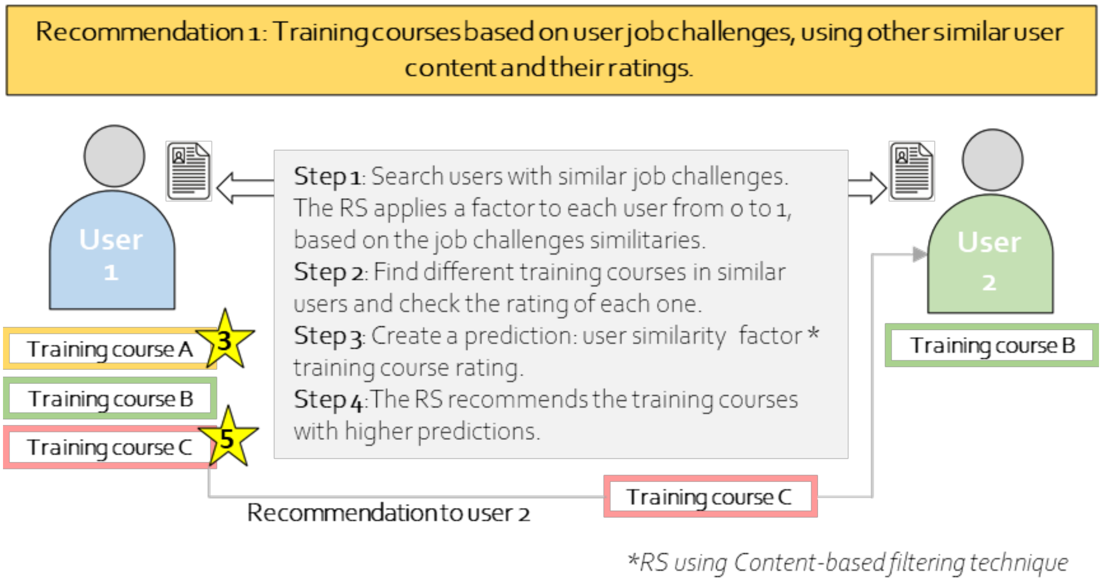


Figure 17. ALLVIEW RS to People users – recommendation 1

Figure *18* shows the process to recommend training courses to People users based on previous training courses already done by the user. It has been developed using the content-based filtering technique. The key in this developed algorithm is to compare key fields existing on training courses included on the user profile (category, language, partner and keywords fields) and those key fields existing on new training courses. It is also important to consider the ratings of the previous training courses in order to discard training courses with similar features, for example, training courses performed by the same partner.

Additionally, in this case, the algorithm does not compare other users, it is only focused on the previous user experience. Taking as example the Figure *18*, in which the user has already performed the training courses A, B and C, the ALLVIEW RS recommends the training course E. In this example, the system has found key fields on the training course E which can be interesting for the user, considering also the training courses A, B and C ratings.

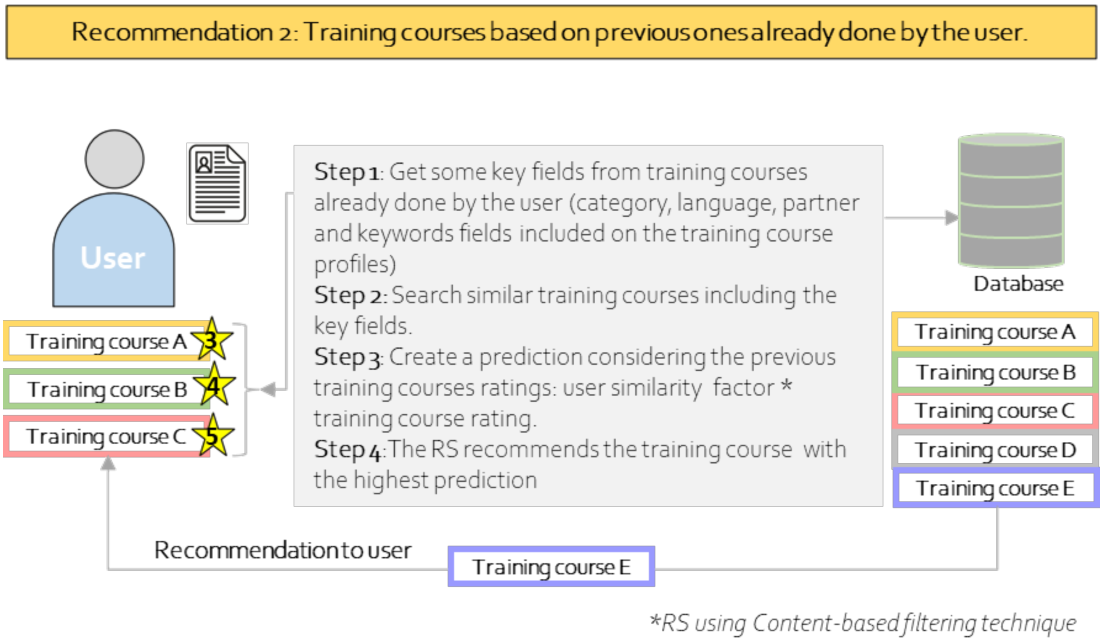


Figure 18. ALLVIEW RS to People users – recommendation 2

The third recommendation is about training courses based on rankings made by other users. In this case, the ML technique applied is the collaborative filtering and the aim is to recommend common training courses based on positive ratings. Taking as example the Figure *19*, the first step is to search users with common training courses with a positive mark. Once the ALVIEW RS has selected 2 users, the user 1 and the user 2 in Figure *19*, the search is focused on other training courses with a positive mark. In this example, the training course A can be recommended to the user 2, and the training course D can be recommended to the user 1.

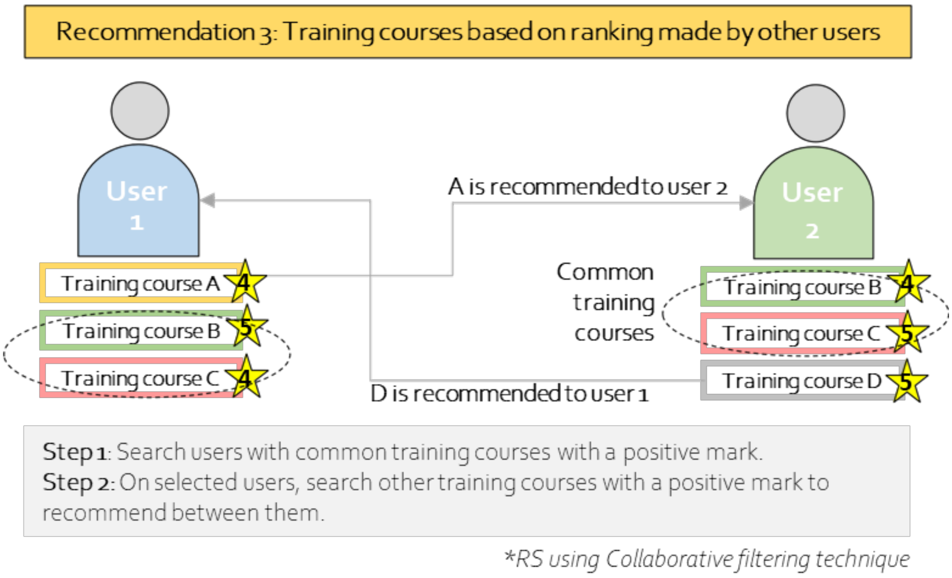


Figure 19. ALLVIEW RS to People users – recommendation 3

Not all users fill in their job challenges and rate training courses in case they perform someone, but recommendations should be implemented for this type of users too. Recommendation 4 in Table 2 uses the content-based filtering technique to recommend training courses based on the skills required by Company users on their job training offers. The steps followed are shown in Figure 20. Firstly, the ALLVIEW RS gets skills from job positions offers published by Company users. Next, it searches training courses to improve or learn those required skills. Finally, the ALLVIEW RS recommends the obtained training courses to users without job challenges/training courses ratings.

Figure 20. ALLVIEW RS to People users – recommendation 4

Apart from training courses, the ALLVIEW RS makes also recommendations about job position offers. In this type of recommendations, the system uses the user profile skills, and the job position offers included in the database by Company users. The recommendation is made based on the similarities between the skills required by the job position offers and the user skills. Figure *21* shows the steps to follow using the content-based filtering technique. Firstly, the system gets the skills that the user has included in the user profile. Next, those skills are compared with every job position offer. Finally, the ALLVIEW RS recommend the most appropriate four job position offers to the user.

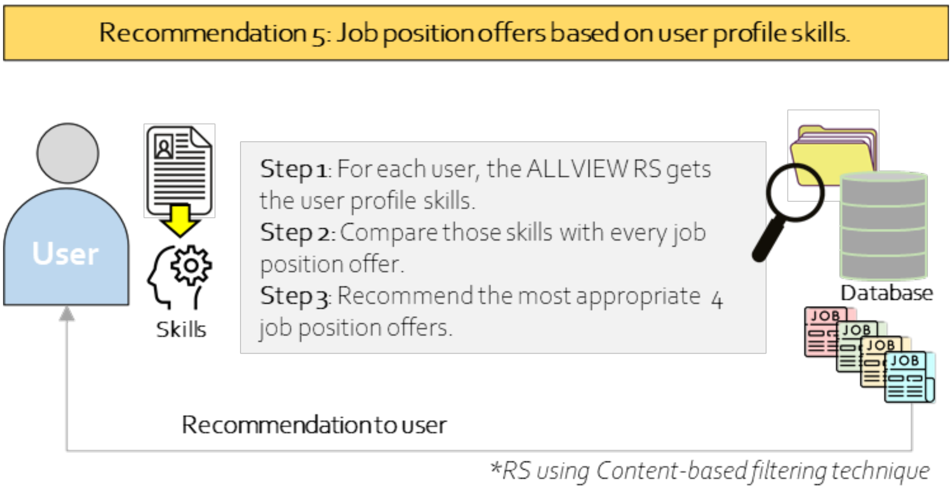


Figure 21. ALLVIEW RS to People users – recommendation 5

The recommendation number 6 in the Table *2* does not implement ML, but it is important for this type of application too. Learning paths are the pathways that a user could follow to achieve new learning challenge and make easier the option to apply to new job position offers. Considering this aim and the example in Figure *22*, the steps followed by the ALLVIEW platform are: 1) get the user job challenges (e.g. occupation B), 2) get the current user skills from the user profile (e.g. skill 3 and skill 4), 3) obtain the essential and optional skills for the user job challenges from ESCO classification (occupation B requires 4 skills), and 4) the platform recommend skills the missing skills in the current user profile to achieve that job challenges (the user needs to learn the skills 1 and 2 to achieve the occupation B).

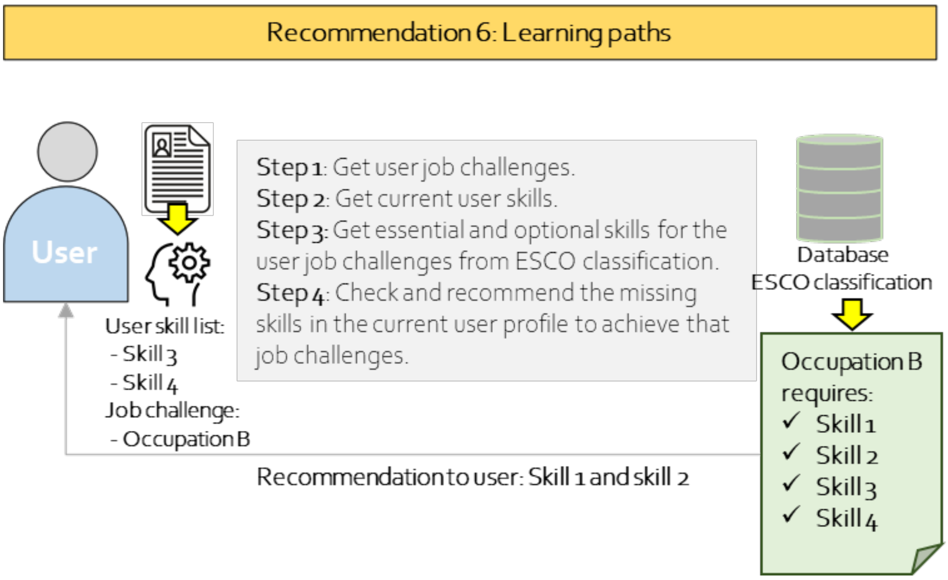


Figure 22. ALLVIEW RS to People users – recommendation 6

3.2 Making recommendations to Training Provider users

There are two types of recommendations for Training Provider users. One includes the recommended candidates for Training Provider users' training courses based on keywords from job challenges, using two ML techniques: content-based filtering and automated keyword extraction. As Figure 23 shows, the ALLVIEW RS obtains and compares keywords from user job challenges and from training courses. The recommended candidates are those who have similar keywords offered in the training course. In the example of the Figure 23, the orange user is not recommended because the keywords got from the job challenge is not like the training course keywords.

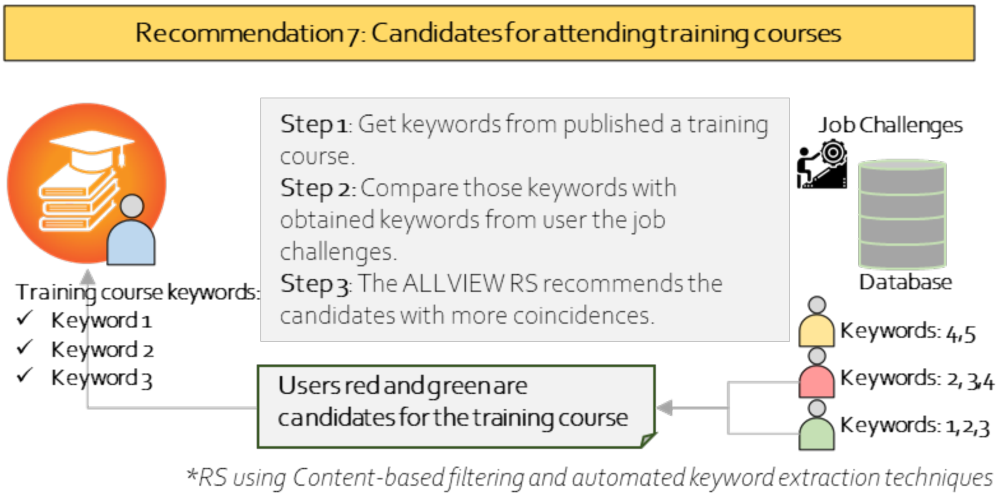


Figure 23. ALLVIEW RS to Training Provider users – recommendation 7

On the other hand, the second recommendation is about the demanded skills by companies that help Training Provider users to prepare new training courses (see Figure 24). In this case, the algorithm does not use ML techniques, but it analyses the job position offers looking for the most popular demanded skills.

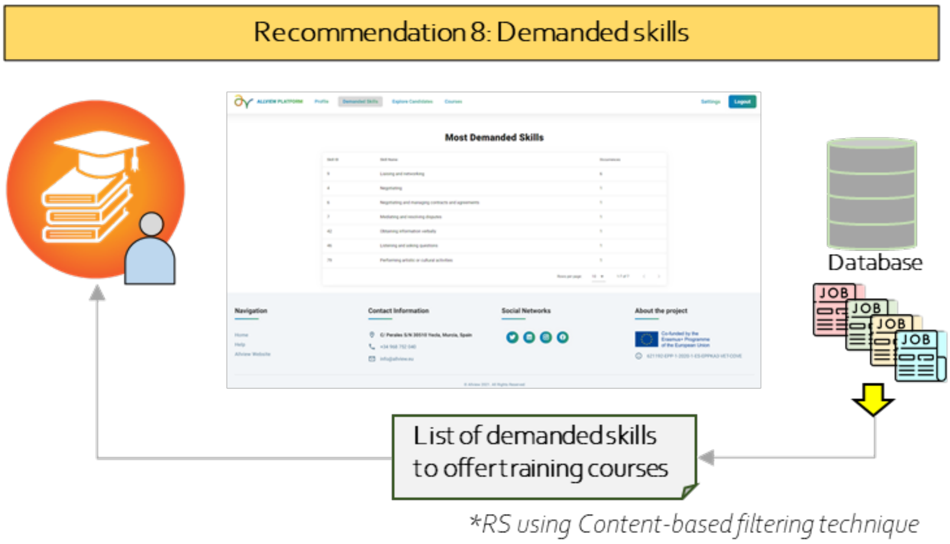


Figure 24. ALLVIEW RS to Training Provider users – recommendation 8

3.3 Making recommendations to Company users

There are 3 recommendations implemented for Company users, 2 of them are similar to other People user recommendations. The recommendation with id 9 implements the same algorithm that the recommendations with id 2 (see Table 2). In this case, the ALLVIEW RS recommends training courses for Companies' employees based on previous training courses done (see Figure 25). The recommendation with id 10, where the ALLVIEW RS recommends training courses based on ranking made by other Company users (see Figure 26), implements the same algorithm that the recommendations with id 3. The third recommendation uses the ML content-based technique to recommend candidates for job position offers based on the required skills. Figure 27 shows the steps of this algorithm, in which the ALLVIEW RS gets the required skills from the job position offers and compare the information with the skills of each People user in order to recommend the appropriate candidates.

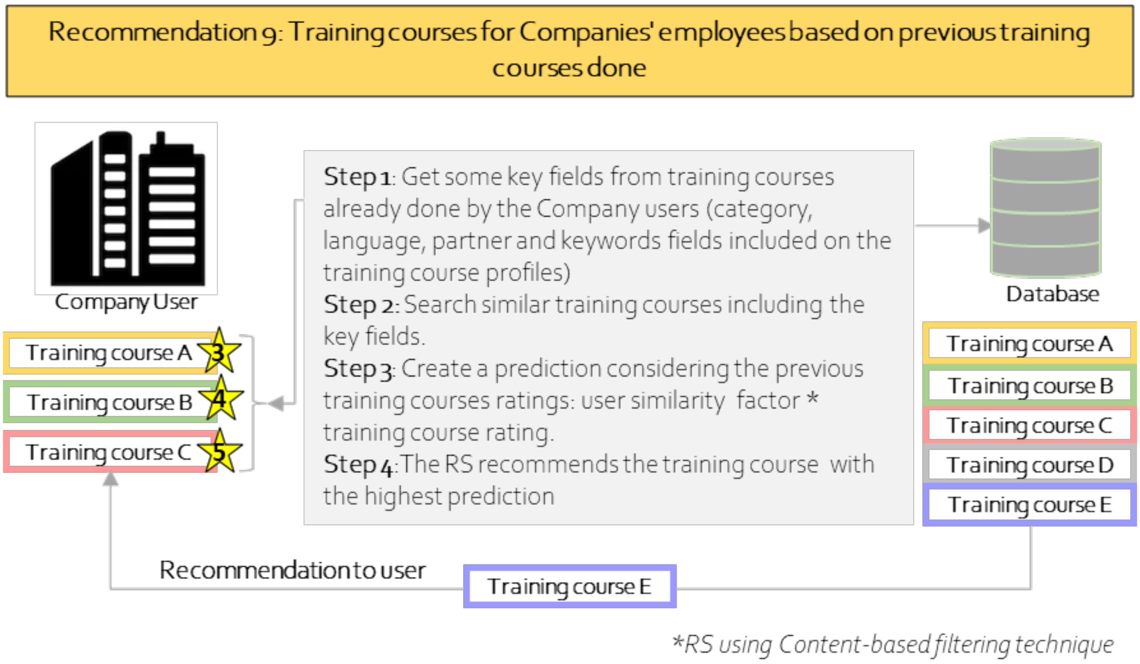


Figure 25. ALLVIEW RS to Company users – recommendation 9

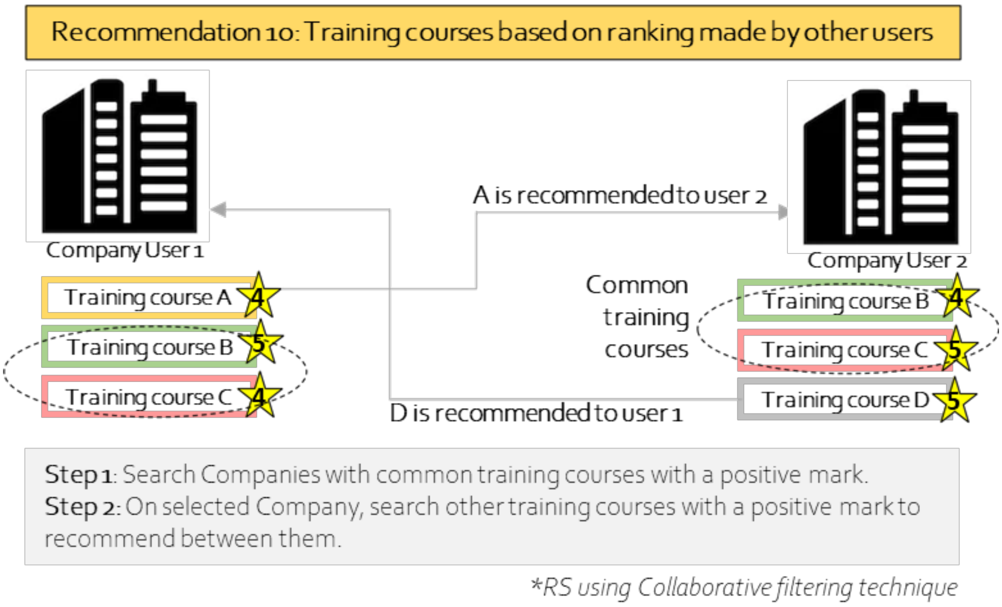


Figure 26. ALLVIEW RS to Company users – recommendation 10

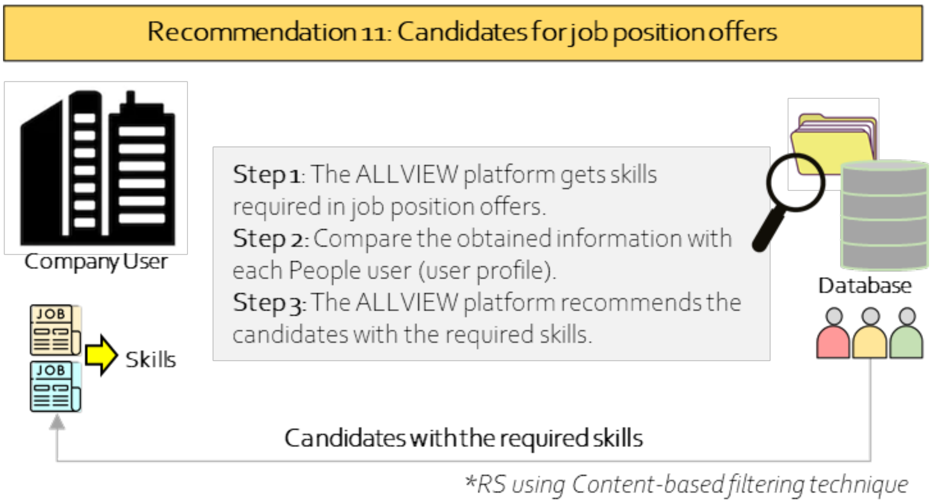


Figure 27. ALLVIEW RS to Company users – recommendation 11

3.4 Common recommendations

Finally, there is a recommendation that is not based on ML techniques, but it uses the IMDb weighted average method to recommend the top training courses. IMDb publishes weighted vote averages rather than raw data averages. It is not the same that a course has 10 ratings with a weighted rating 5 stars, and a course with 100 ratings with a weighted rating of 4 stars. The equation for the weighted rating (WR) is the following:

$$WR = (v \div (v+m)) \times R + (m \div (v+m)) \times C$$

Where:

R = average rating
v = number of votes for the training course = (votes)
m = minimum votes required to be listed in the Top
C = the mean vote across the whole data

Allview

4

Evaluation of ML techniques with data sets – training-validation-output

4. Evaluation of ML techniques with data sets – training-validation-output

The evaluation of the developed ML techniques is essential to determinate that the algorithms have been built correctly. To evaluate the methods described in previous sections, a training set and the expected output have been defined for each use case defined in Table 2. The training dataset is the initial data used to train ML models. Training datasets are fed by ML algorithms to teach them how to make predictions or perform a desired task. Below, the use cases have defined to check that the implementation of the ML methods show good results. Some examples are created with specific characteristics according to the definition.

Use case 1: To recommend training courses based on user job challenges, using other similar user content and their ratings.

- User type: People.
- Training data set: Select at least three users with similar job challenges with two of these users, rate some courses with a good mark.

To perform use case 1 test, three user profiles have been created with the data showed in Table 3, Table 4 and Table 5. All of them have the same job challenge: project manager, and users 1 and 2 have rated some courses with a good rating (see Table 6).

Table 3: User 1 profile

User 1 profile		
Personal information		User 1, people1@test.com, Murcia, Spain, 666666666
Curriculum	Skills	Manage project information, train employees, working in teams
	Competences	Adapt to changing situations, check the production schedule
	Job Experiences	Research engineer
	Educations	Telecommunications
Job challenges		Project manager

Table 4: User 2 profile

User 2 profile		
Personal information		User 2, people2@test.com, Roma, Italy, 555555555
Curriculum	Skills	Coordinating activities with other, train employees, manage several projects

	Competences	Implement data quality processes, meet commitments
	Job Experiences	Sales engineer
	Educations	Telecommunications
Job challenges		Project manager

Table 5: User 3 profile

User 3 profile		
Personal information		User 3, people3@test.com, Brussels, Belgium, 222222222
Curriculum	Skills	Manage tests, record test data, analyse test data
	Competences	Provide customer follow-up services
	Job Experiences	Test engineer
	Educations	Telecommunications
Job challenges		Project manager

Table 6: Rated training courses for use case 1.

Rated training courses		
Training course	Rated by	Average rating
Project Management Principles and Practices	User 1, User 2	4
Fundamentals of Project Planning and Management	User 1, User 2	5
Computer course	User 1, User 2	4
Augmented reality as infrastructure for improvement of communication in construction projects	User 1, User 2	5

- Expected output: The system should recommend to the third user, the training courses that similar users (users with the same job challenges) have rated positively (the training courses included in Table 6).
- Results: Figure *28* shows the recommendations made to the user 3. The results match with the expected output.

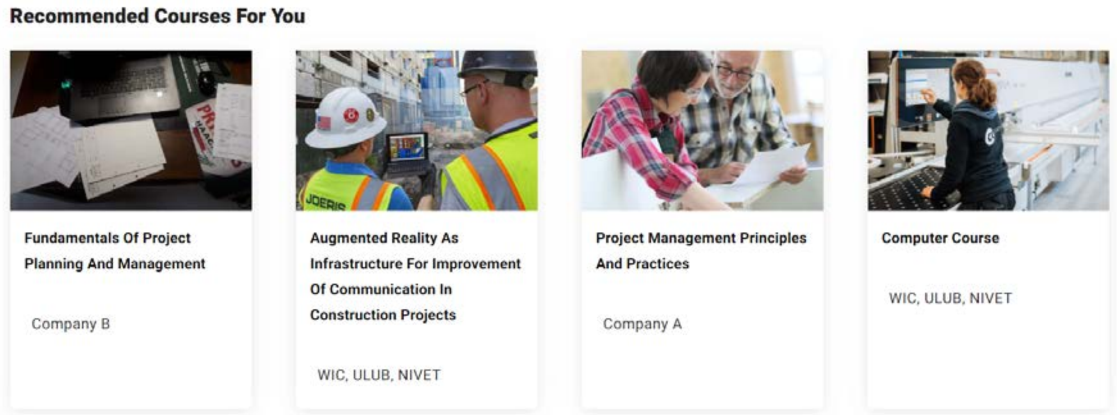


Figure 28. Use case 1 results

Use case 2: To recommend training courses based on previous courses already done by the user.

- User type: People.
- Training data set: Select a user (the User 1 defined in Table 3) and rate with a high mark a training course with similar characteristics.

The User 1 has rated with a high mark the course titled "Project Management Principles and Practices" in Table 7. This tables show other training courses that could be interesting for the User 1.

Table 7: Training courses details

Training courses (required fields)				
Title	Engineering project management	Construction management	Product manager	Project Management Principles and Practices
Company name	Company A	Company B	Company B	Company A
Category	Management skills	Management skills	Management skills	Management skills
Keywords	Management, engineering projects	Management	Management	Management
Language	English	English	English	English

- Expected output: The system should recommend to this user training courses with shared characteristics from the training courses that the user has rated positively. In this example, the training courses included in Table 7 should be recommended to User 1.

- Results: Figure 29 shows the recommendations made to the user 1. The results match with the expected output.

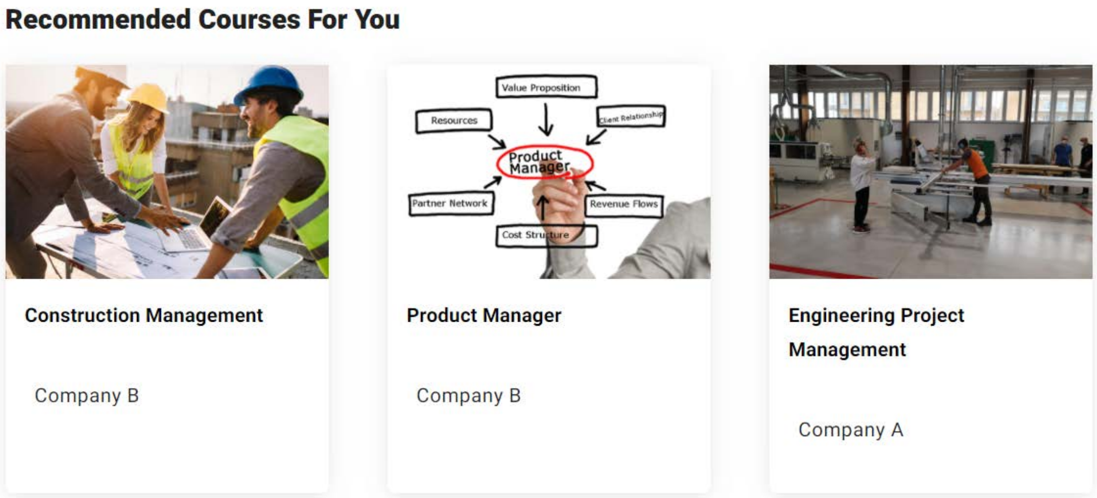


Figure 29. Use case 2 results

Use case 3: To recommend training courses based on ranking made by other users with the same interests.

- User type: People.
- Training data set: Select three users (User 1, User 2 and User 3 already defined above) and rate four courses in total (see Table 8) with a high mark but with "holes". For instance: User 1 rates the training courses 1, 2 and 3; the User 2 rates the training courses 1, 2 and 4; and the User 3 rates the training courses 1, 3 and 4.

Table 8: Rated training courses for use case 3

Rated training courses			
Id	Training course	Rated by	Rating
1	Project Management Principles and Practices	User 1, User 2, User 3	4
2	Fundamentals of Project Planning and Management	User 1, User 2	5
3	Engineering Project Management	User 1, User 3	4
4	Product manager	User 2, User 3	4

- Expected output: The system must recommend with a high score to the three users the course that they have not rated.
- Results: Figure 30, Figure 31 and Figure 32 show the recommendations made to the User 1, User 2 and User 3 respectively. The results match with the expected output.

Figure 30. Results for user 1

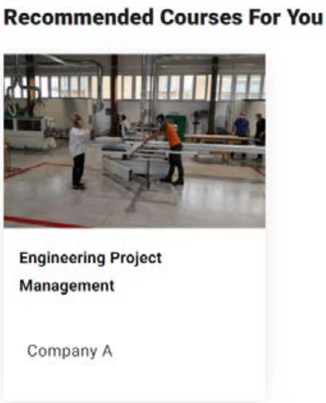


Figure 31. Results for user 2

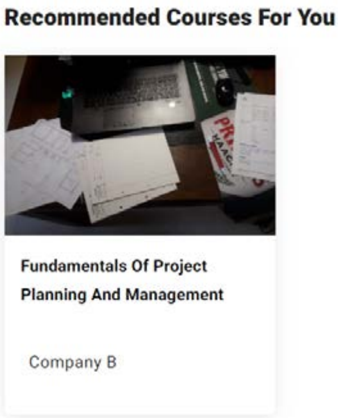


Figure 32. Results for user 3

Use case 4: To recommend training courses demanded by companies, considering the skills included in job position offers. This recommendation is made to People users who do not have included any job challenge or ranked any training course.

- User type: People.
- Training data set: Create a user with no user-profile and no ratings (User 4). Also, some job position offers must be available in the platform (see examples in Table *10*).

The system calculates keywords from job position offers and compares with the keywords of each training course. The calculated keywords from job position offers included in Table 10 are: 'administer', 'complete', 'projects', 'people', 'forces', 'profitability', 'relevant', 'tests', 'coherence', 'leveraging'. Note that the system uses ESCO classification, and every professional sector and skill has associated some descriptions that the system use to get the keywords. To test this, two training courses have been modified adding the following keywords in Table *9*.

Table 9: Training courses keywords

Training course	Keywords
Project Management Principles and Practices	Projects, manage tests, administer, complete, people, leveraging
Engineering Project Management	Management, projects, test, people, administer

Table 10: Job position offers

Job position offers (some required fields)			
Title	Management assistant	Test Engineer	Project manager

Professional sector	Engineering professionals	Engineering professionals	Engineering professionals
Essential Skills	Manage tests	Manage tests	Manage several projects
Main language	English	English	English
Type of contract	Practice	Temporal	Undefined

- Expected output: The user must receive recommendations of courses from the most demanded skills from the job offers.
- Results: Figure 33 shows the recommendations made to the user 4. The results match with the expected output.

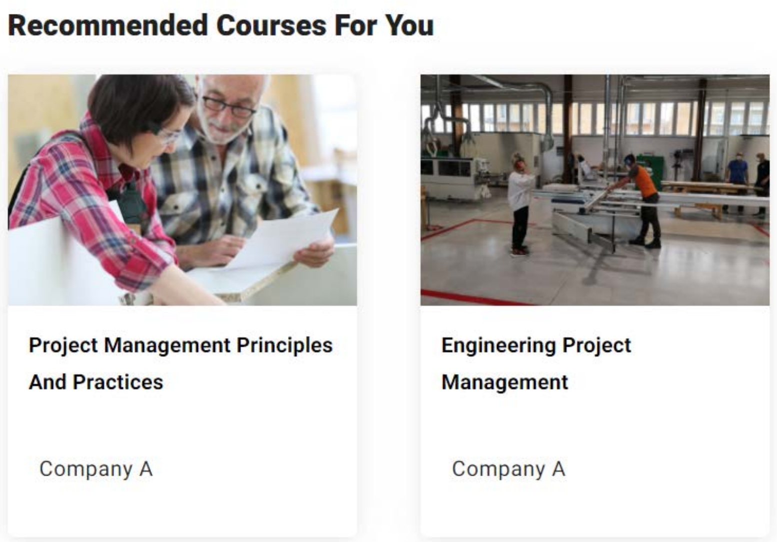


Figure 33. Use case 4 results

Use case 5: To recommend job position offers based on user profile skills. Ranking is not used.

- User type: People.
- Training data set: Create a user and select some skills. Create a job with similar skills to the user.

The data used for this example is a user with the skill "Manage several projects" in the profile, and some job position offers that require the same skill. The job position offers are titled "Management Assistant", "Project Manager", "Office Manager" and "Department Manager".

- Expected output: The user must receive recommendations about jobs with similar skills required.
- Results: Figure 34 shows the recommendations made to the user. The results match with the expected output.

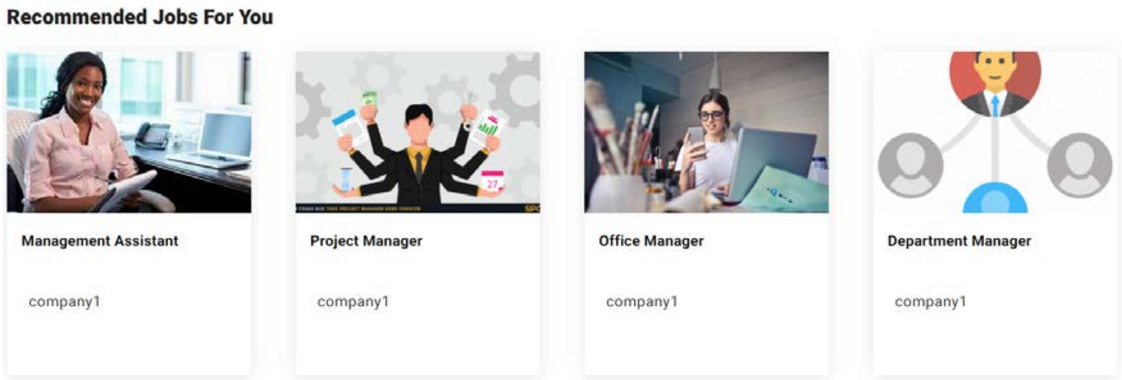


Figure 34. Use case 5 results

Use case 6: Learning paths based on user challenges and current user competences.

- User type: People.
- Training data set: A user (the User 1 already defined) with some job challenges (project manager in this example) and with some skills (Manage project information, train employees, working in teams in Table 3).

The system calculates the essential and optional skills to achieve the user job challenge and compare with the user profile skills.

- Expected output: Job challenges page must display the required and optional skills needed to achieve that job challenge.
- Results: Figure 35 shows the recommendations made to the user. The symbol means that the user already has this skill, and the symbol means that the user does not have this skill to achieve the job challenge. The results match with the expected output.

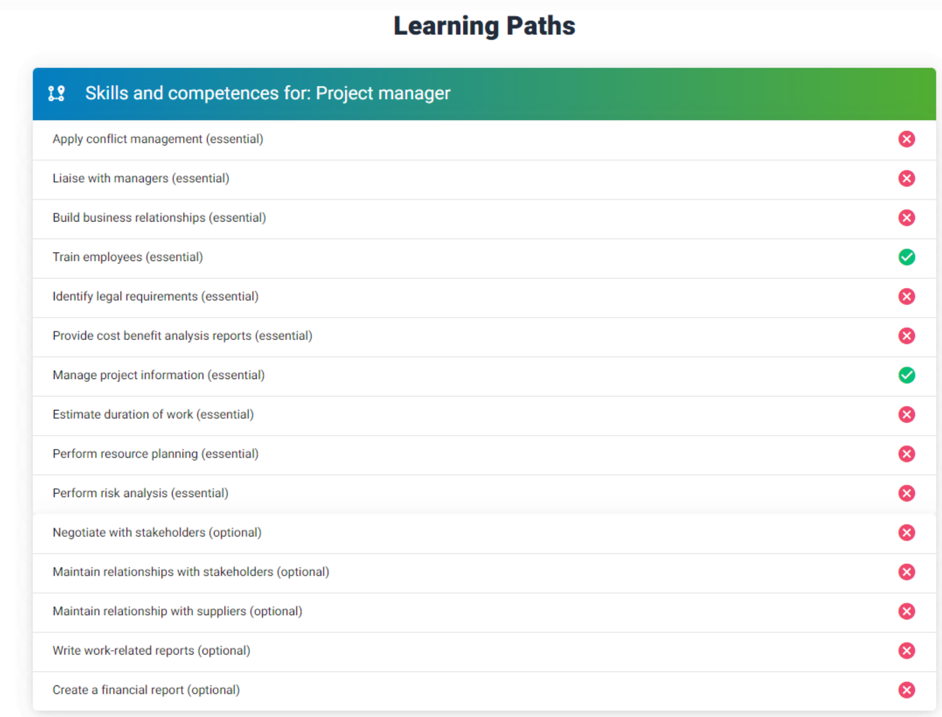


Figure 35. Use case 6 results

Use case 7: To recommend candidates for attending training courses based on keywords from job challenges.

- User type: Training provider.
- Training data set: Some People users with similar job challenges and create a new training course focused on that job challenge with some keywords.

The users to test this use case are defined in Table 3, Table 4and Table 5, in which the user job challenge is "Project manager". The system calculates the keywords from this user job challenge and compare with the keywords of each training course. In this example, the user training provider tested offers a training course with similar keywords that the user job challenge ('project', 'ensure', 'satisfied', 'completed', 'budget').

- Expected output: The system must recommend candidates 1, 2 and 3, who have similar keywords extracted, from the occupation's descriptions, to the training provider that offers a training course with similar keywords.
- Results: Figure 36 shows the recommendations made to the user. The results match with the expected output.

Figure 36. Use case 7 results

Use case 8: To recommend demanded skills to prepare new training courses.

- User type: Training provider.
- Training data set: Some jobs with the same required skills.
- Expected output: The training provider must see these most demanded skills.
- Results: Figure 37 shows the recommendations made to the user. The results match with the expected output.

Most Demanded Skills

Skill Name	Occurrences
Manage several projects	4
Manage tests	2
Facilitate official agreement	1

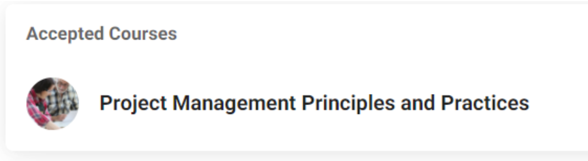
Rows per page: 10 1-3 of 3

Figure 37. Use case 8 results

Use case 9: To recommend training courses for Companies' employees based on previous training courses done.

- User type: Company.
- Training data set: Select a user and rate with a high mark some courses with similar characteristics.

This use case is similar to use case 2. In this example, the user Company 1 has performed and rated the training course “Project Management Principles and Practices” with a high mark, and then, the system will look for similar training courses.



- Expected output: The system must recommend to this user courses with shared characteristics from the courses that the user has rated positively.
- Results: Figure 38 shows the recommendations made to the user. The results match with the expected output.

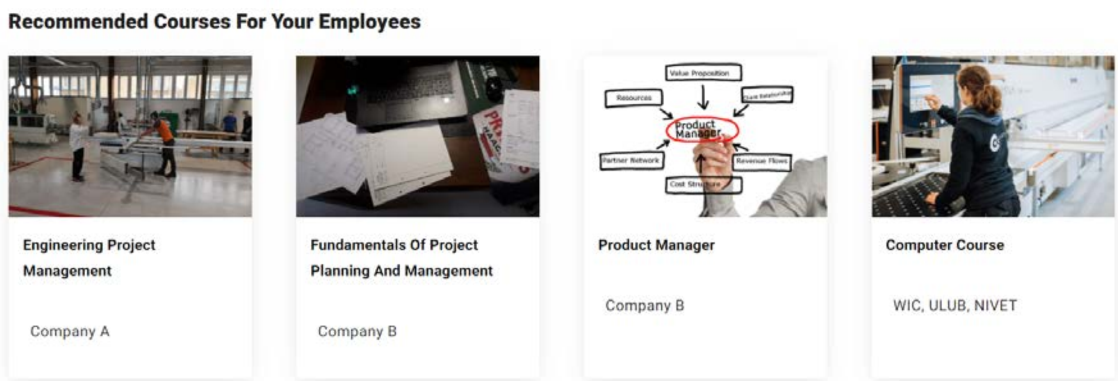


Figure 38. Use case 9 results

Use case 10: Training courses for Companies' employees based on other users with the same interests.

- User type: Company.

Training data set: Select three users and rate four courses in total with a high mark but with “holes”. This use case is like use case 3. The training courses rated are included in Table *8*.

- Expected output: The system must recommend with a high score to the three users the course that they have not rated.
- Results: Figure 39, Figure 40 and Figure 41 show the recommendations made to the users. The results match with the expected output.

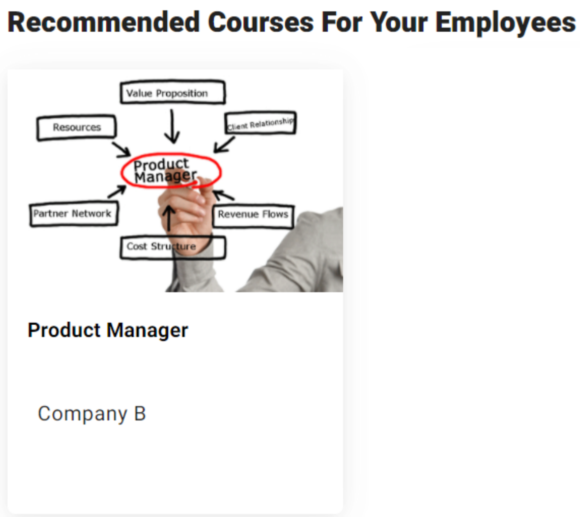


Figure 39. Use case 10 result for Company 1 user

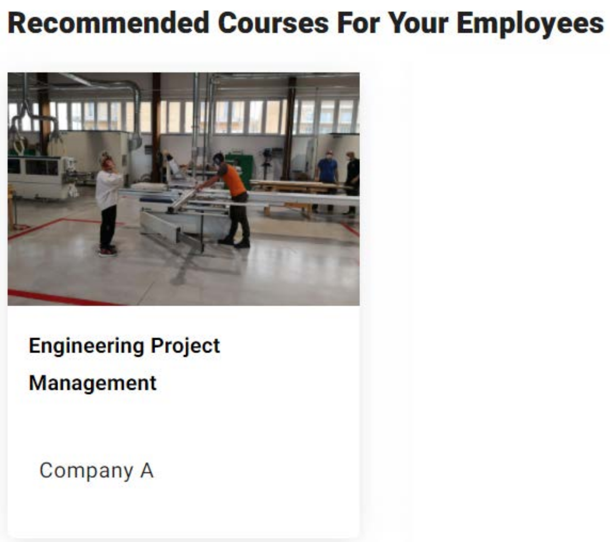


Figure 40. Use case 10 result for Company 2 user

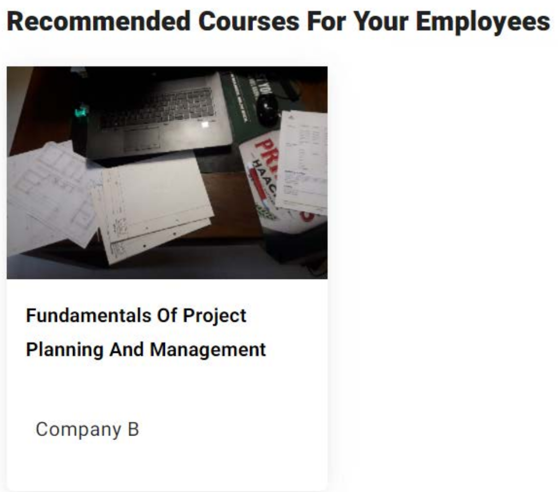


Figure 41. Use case 10 result for Company 3 user

Use case 11: To recommend candidates for job position offers (based on skills required in job position offers).

- User type: Company.
- Training data set: Some users with some skills and create a job offer with the same skills.

The user 2, 3, 4 and 5 include in their user profile the skill "Manage several projects" as we can see in Figure 42.

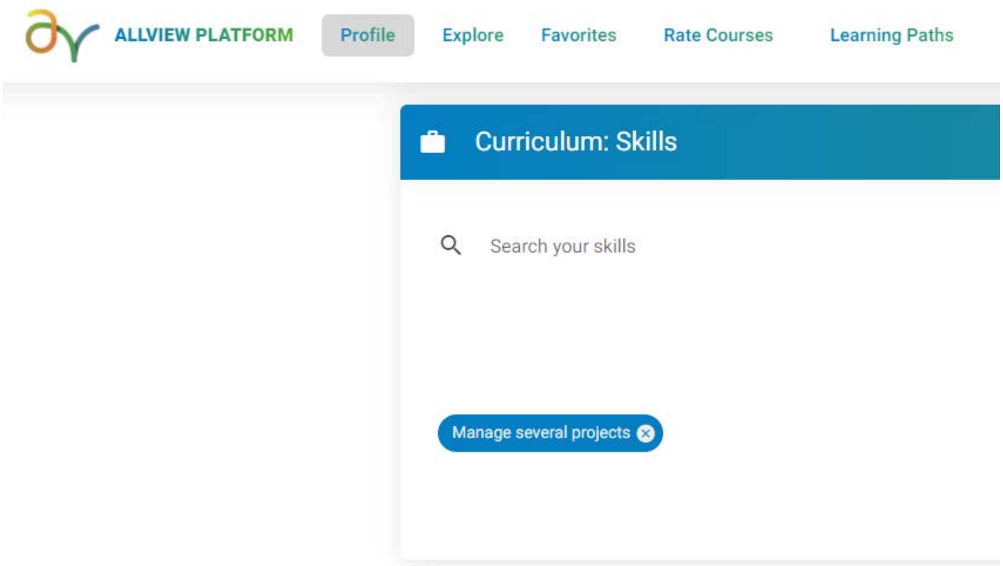


Figure 42. Use case 11 – User profile: skills

Next, the job position offer published by the user Company 1 is titled "Office manager". As Figure 43 shows, the essential skill defined for this job position offer is "Manage several projects".

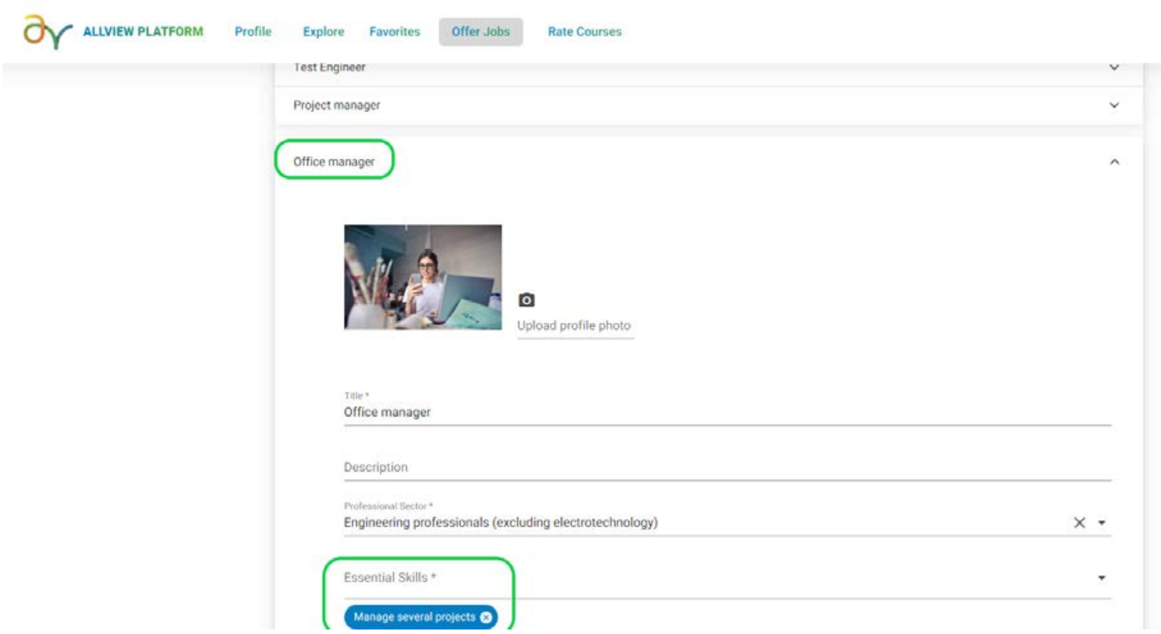


Figure 43. Use case 11 – Job position offer

- Expected output: The system must recommend these 4 users as candidates for the job position offer published by the Company user.
- Results: Figure 44 shows the recommendations about candidates made to the user Company 1. The results match with the expected output.

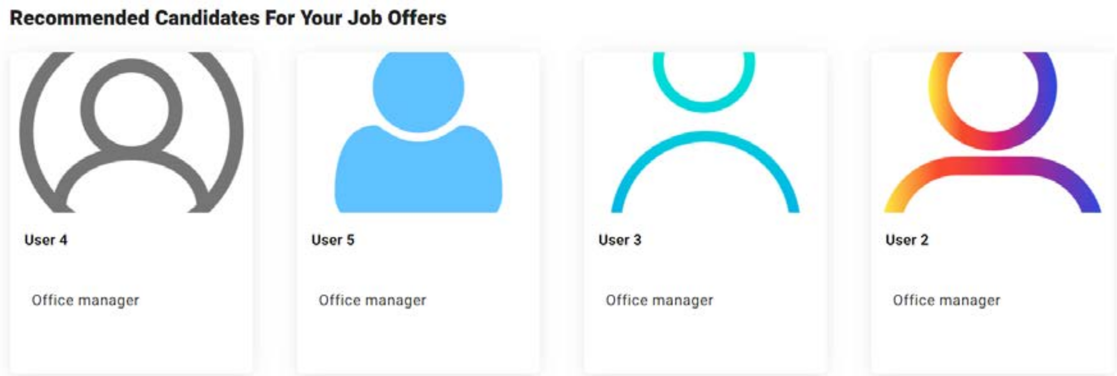


Figure 44. Use case 11 results

Use case 12: Top rated training courses.

- User type: People, Training provider and Company.
- Training data set: Create a training course with many ratings and a rating average of 4. Create a training course with few ratings (3 o + more than three) and a rating average of 5.

Two tests have been planned. In the first test, 3 users have rated with 4 the training course "Computer course" and with 5 the training course "Engineering Project Management" (see Accepted Rating in Figure 45). In the second place, 7 additional users have rated the training course "Computer course" with 4 (see Accepted Rating in Figure 46).

- Expected output: On the first test, in which the quantity of ratings is the same for both training courses, the training course with a higher score must be showed in a higher position. On the second test, the training course with many ratings and normal score ("Computer course" in this example) must show with a higher ranking than the course with few ratings and high score ("Engineering Project Management" in this example).
- Results: Figure 45 and Figure 46 show the top-rated courses. The results match with the expected output.

Accepted Ratings

title	User	Comment	Justification	Rating
Engineering Project Management	User 4			5
Computer course	User 4			4
Engineering Project Management	User 3			5
Computer course	User 3			4
Engineering Project Management	User 5			5
Computer course	User 5			4

Rows per page: 10 1-6 of 6

Top Rated Courses

Engineering Project Management

Company A

Computer Course

WIC, ULUB, NIVET

Figure 45. Use case 12 results (1)

Accepted Ratings

title	User	Comment	Justification	Rating
Engineering Project Management	User 4			5
Computer course	User 4			4
Engineering Project Management	User 3			5
Computer course	User 3			4
Engineering Project Management	User 5			5
Computer course	User 5			4
Computer course	User 1			4
Computer course	User 2			4
Computer course	company1			4
Computer course	company2			4

Rows per page: 10 1-10 of 13

Accepted Ratings

title	User	Comment	Justification	Rating
Computer course	company3			4
Computer course	company4			4
Computer course	company5			4

Rows per page: 10 11-13 of 13

Figure 46. Use case 12 results (2)

5

Final deployment and public repository

5. Final deployment and public repository

The initial and debugging tests were planned in a UPCT server, being this the repository and framework for the solution development. Once the platform has been debugged, it has been launched in a dedicated server in the Cloud. Not only because the UPCT server has technical limitations, but also because The Cloud solutions have advantages like 24/7 support, GPDR, security, data privacy, etc. To hire a Cloud solution there are several options. Table 11 shows some of them with the main features that these services offer. Currently, ALLVIEW platform only requires two services: backup and virtual machine, but others are also considered, thinking in future versions platform. The selected Cloud solution provider is DigitalOcean 12. DigitalOcean provides developers cloud services that help to deploy and scale applications that run simultaneously on multiple computers. It is similar to other Cloud solutions, but it has been selected because the ALLVIEW developer technical team already know this platform and cover the needs.

Table 11. Cloud solutions providers

	aws	VULTR	DigitalOcean	Microsoft Azure	
Server provider	**Amazon**	**Vultr**	**DigitalOcean**	**Azure**	**Google Cloud**
Cloud storage	✓	✓	✓	✓	✓
Managed database	✓	✓	✓	✓	✓
Backup	✓	✓	✓	✓	✓
Virtual machine	✓	✓	✓	✓	✓
Url	https://aws.amazon.com/	https://www.vultr.com/	https://www.digitalocean.com/	https://azure.microsoft.com/	https://cloud.google.com/

Once the cloud solution provider has been selected, the final deployment showed in Figure 47 can be described. The production server hired in DigitalOcean platform is an Ubuntu server 20.04 with 8 Intel vCPUs, 16 GB RAM, 320 GB SDD storage, >250 Mbps network bandwidth and authentication with SSH keys. The MySQL database can be managed by the admin through the application phpMyAdmin (more details about the database are included in deliverable D1.3).

On the other hand, the open-source code of the ALLVIEW platform has been allocated on a public repository, fulfilling the commitment in the project description. During the development, the source code was configured as private, but once the platform is finished and ready to be used, the

12 https://www.digitalocean.com/

platform will be configured as public. The public repository selected was Bitbucket 13. Bitbucket is a Git-based source code repository hosting service owned by Atlassian. Bitbucket offers both commercial plans and free accounts with an unlimited number of private repositories.

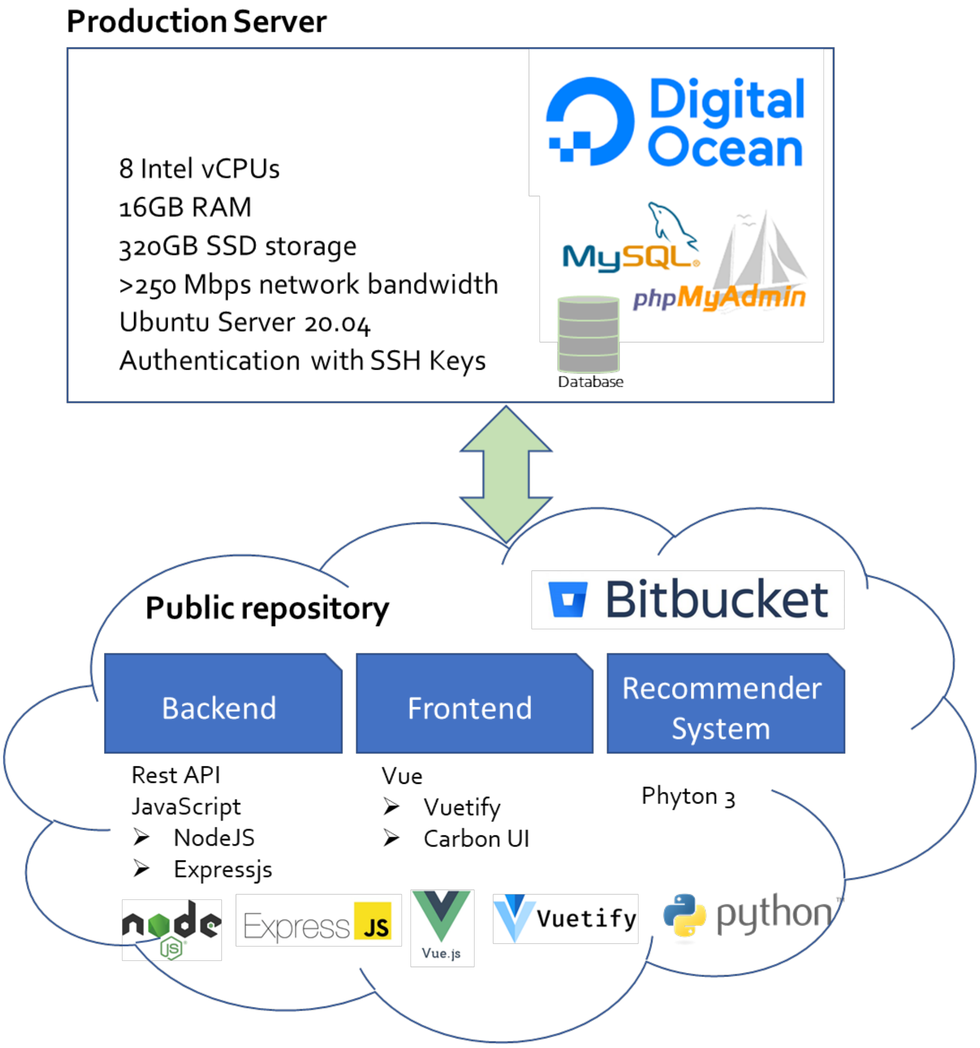


Figure 47. Final deployment

The project is composed of three repositories (see Figure 47 and Figure 48): (1) allview-backend, (2) allview-frontend, and (3) allview-recommender-system. The repository allview-backend (see Figure 49) contains the REST API of the ALLVIEW Platform, written in JavaScript using Nodejs and Expressjs. The repository allview-frontend (see Figure 50) contains the Frontend of the ALLVIEW Platform, written in Vue using Vuetify and CarbonUI. The repository allview-recommender-system (see Figure 51) contains the recommender system of the ALLVIEW Platform, written in Python 3. Note that all these technologies were introduced in depth in Deliverable 1.2.

Each repository on Bitbucket includes a file README with the structure of that repository, a project overview, the software dependences that must be installed, the configuration, how to run the service and how to compile for production.

13 *https://bitbucket.org/*

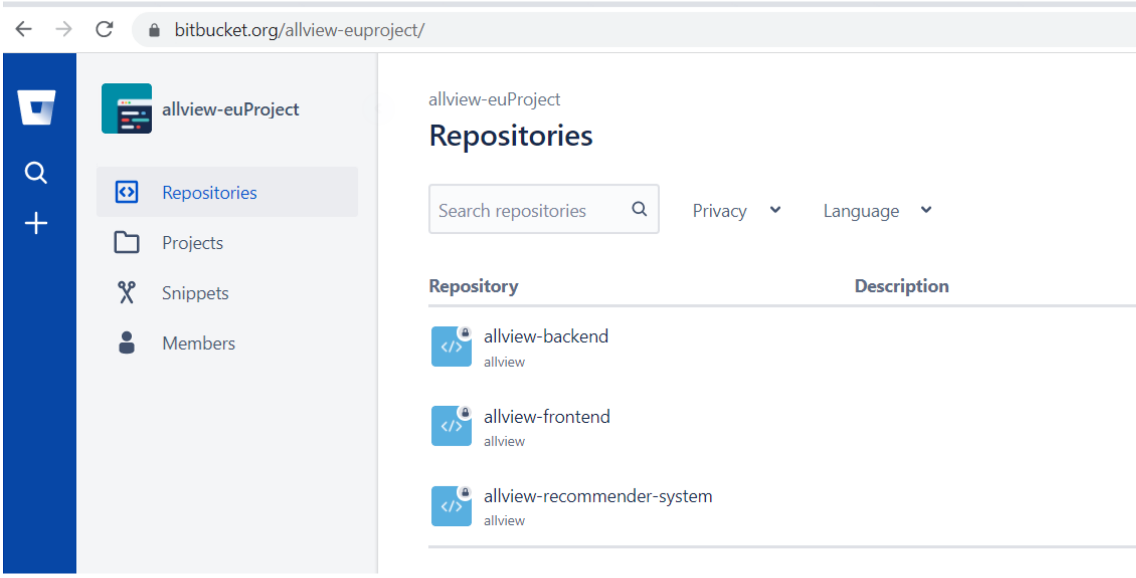


Figure 48. Public repository

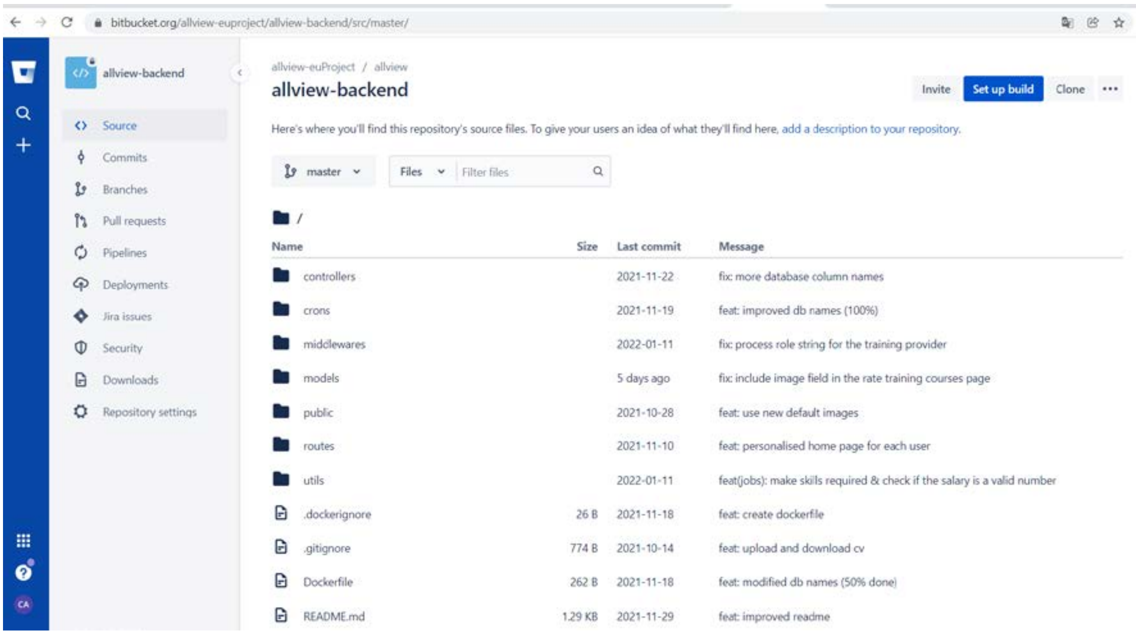


Figure 49. Repository Allview-backend

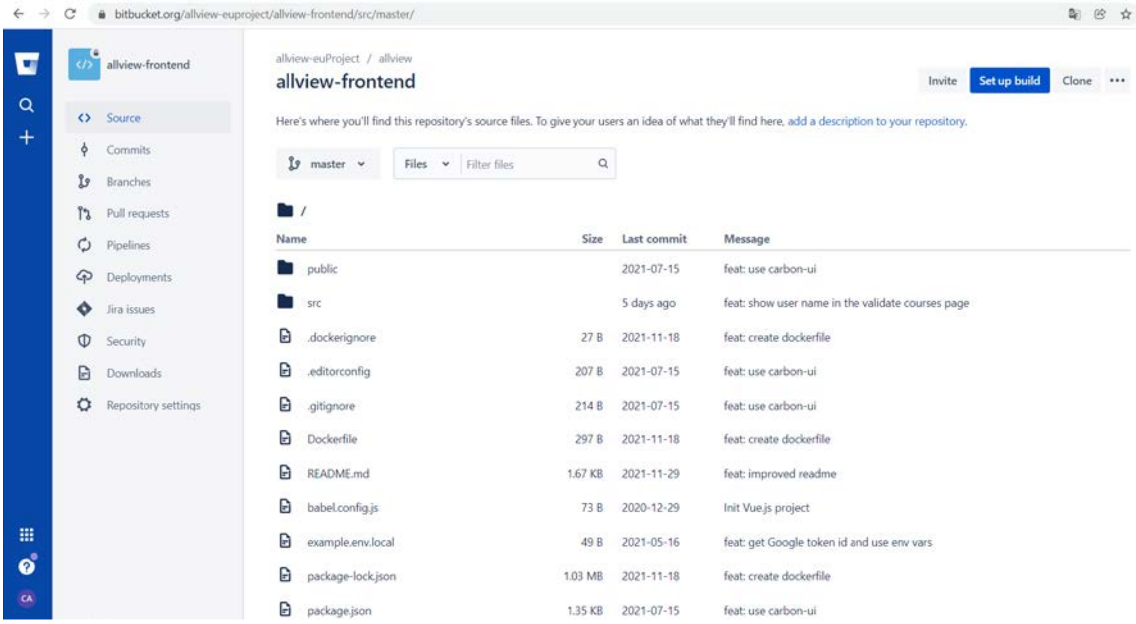


Figure 50. Repository Allview-frontend

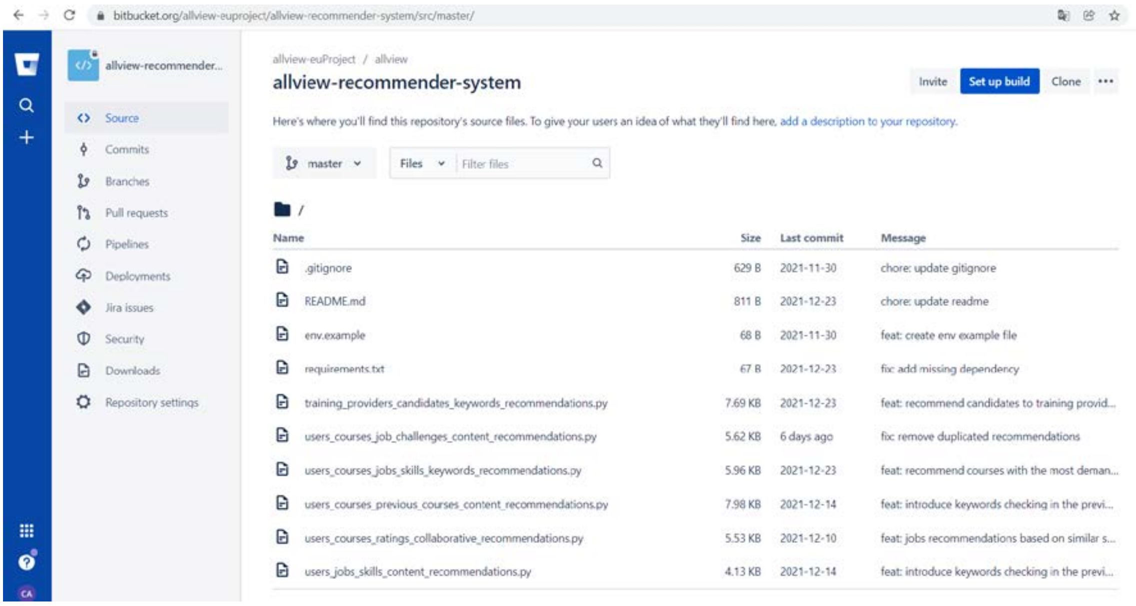


Figure 51. Repository Allview-recommender-system

6

Conclusions

6. Conclusions

During the development of task 1.4, the main categories of AI/ML algorithms were analyzed, identifying the most appropriated to be implemented in the ALLVIEW platform. There are a vast quantity of AI/ML algorithms, but the key is on selecting the best depending on the platform requirements. In this project, the platform that is being developed must include recommendation about training courses, job positions offers and even candidates to those training courses and job position offers. Based on that, RS classified as unsupervised learning algorithms have been developed, including contend based and collaborative filtering techniques.

Moreover, the identification of ML techniques has been performed. Each type of recommendation has been linked to a ML technique, also identifying the type of user, the recommendation description, and an example of each one (remember Table 2). An algorithm has been defined and described in detail for every recommendation, including a diagram in this report to facilitate understanding. To sum up, the recommendations offered by the ALLVIEW platform by type of user are:

- ✓ To People users:
  - Training courses based on user job challenges, using other similar user content and their ratings.
  - Training courses based on previous courses already done by the user.
  - Training courses based on ranking made by other users with the same interests.
  - Training courses demanded by companies, considering the skills included in job offers. This recommendation is made to People users who do not have included any job challenge or ranked any training course.
  - Job position offers based on user profile skills. Ranking is not used.
  - Learning paths based on user challenges and current user competences.

- ✓ To Training Provider users:
  - Candidates for attending training courses based on keywords from job challenges.
  - Demanded skills to prepare new training courses.
  - To Company users:
  - Training courses for Companies' employees based on previous training courses done.
  - Training courses for Companies' employees based on other users with the same interests.
  - Candidate for job position offers (based on skills required in job position offers).
  - To all users:
  - Top rated training courses.

After the definition of the ALLVIEW platform recommendations, an evaluation of the developed RS has been planned. The evaluation of the ML techniques developed is essential before deploying the platform, for that, a training data set and the expected output have been defined for every use case. For each use case the results have been executed and included in this report to show the correct performance of the ALLVIEW platform.

Finally, once the identification of ML techniques has been made and the evaluation of the defined recommendations has been validated, next step has been to plan the final deployment. It is also important to remark that the source code developed has been allocated in a public repository, commitment included in the project description. The public repository selected is Bitbucket

The open-source code developed for the ALLVIEW platform has been organized as follow:

- Repository 1: allview-backend
- REST API written in JavaScript using Nodejs and Expressjs
- Repository 2 allview-frontend
- Written in Vue using Vuetify and CarbonUI
- Repository 3: allview-recommender-system.
- Written in Python 3

About the deployment, the initial testing was performed on a UPCT server, due to technical requirements, but the final deployment was performed on a hired Cloud solution. This report shows a short comparison of the most popular Cloud solutions providers, setting DigitalOceanas the platform selected for ALLVIEW platform.

Allview

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