Sociology Study, May-June 2023, Vol. 13, No. 3, 161-173
doi: 10.17265/2159-5526/2023.03.006

Gamified Learning Systems' Personalized Feedback Report Dashboards via Custom Machine Learning Algorithms and Recommendation Systems

Nymfodora-Maria Raftopoulou
EDISINET S.A., Thessaloniki; Athens University of Economics and Business, Athens, Greece
Petros L. Pallis
EDISINET S.A., Thessaloniki; University of Piraeus, Piraeus, Greece

Gamification in education enables for the holistic optimization of the learning process, empowering learners to ameliorate their digital, cognitive, emotional and social skills, via their active experimentation with game design elements, accompanying pertinent pedagogical objectives of interest. This paper focuses on a cross-platform, innovative, gamified, educational learning system product, funded by the Hellenic Republic Ministry of Development and Investments: *howlearn.* By applying gamification techniques, in 3D virtual environments, within which, learners fulfil STEAM (Science, Technology, Engineering, Arts and Mathematics)-related Experiments (Simulations, Virtual Labs, Interactive Storytelling Scenarios, Decision Making Case Studies), *howlearn* covers learners' subject material, while, simultaneously, functioning, as an Authoring Gamification Tool and as a Game Metrics Repository; users' metrics are being, dynamically, analyzed, through Machine Learning Algorithms. Consequently, the System learns from the data and learners receive Personalized Feedback Report Dashboards of their overall performance, weaknesses, interests and general class competency. A Custom Recommendation System (Collaborative Filtering, Content-Based Filtering) then supplies suggestions, representing the best matches between Experiments and learners, while also focusing on the reinforcement of the learning weaknesses of the latter. Ultimately, by optimizing the Accuracy, Performance and Predictive capability of the Personalized Feedback Report, we provide learners with scientifically valid performance assessments and educational recommendations, thence intensifying sustainable, learner-centered education.

Keywords: gamified education, in-game data analytics, personalized feedback report dashboard, recommendation systems, statistics

Introduction

howlearn is a cross-platform (Windows, Android/iOS and Web) innovative educational learning system product, funded by the Hellenic Republic Ministry of Development and Investments, within which, learners complete three-dimensional, gamified Experiments (gamified three-dimensional educational exercises),

Nymfodora-Maria Raftopoulou, BSc, MA, and MSc, EDISINET S.A., Thessaloniki; Department of Statistics, Athens University of Economics and Business, Athens, Greece.
Petros L. Pallis, BSc, MBA, MSc, and PhD, EDISINET S.A., Thessaloniki; Department of Maritime Studies, University of Piraeus, Piraeus, Greece.

covering their subject material. Through classroom, laboratorial and entrepreneurial simulations, the System's end users (Educational Institutions, Instructors and Learners) are being given the opportunity to fulfil realistic scenarios, related to their educational material, without having to worry about any socio-economic and geographical boundaries, which would imply their exclusion from the learning process, had physical appearance been required (Raftopoulou & Pallis, 2023a).

By making use of a cross-platform game engine, "Unity", *howlearn*'s "Virtual Library of Experiments" supports a wide range of desktop, mobile, console and virtual reality platforms and concentrates (yet is not limited to) on supplying Gamified Education in the area of "STEAM" (Science, Technology, Engineering, Arts and Mathematics), via:

- 10 virtual thematic experimental laboratory simulations (Physics, Chemistry, Biology);
- 5 virtual labs in Mechanics and ICT;
- 10 interactive storytelling (narration) scenarios (English and Mathematics);
- 5 Literature and Arts simulations and case studies;
- 2 educational interactive storytelling & decision-making scenarios, familiarizing learners with the conceptualization of entrepreneurship and innovation.

Additionally, both the Educational Institutions and Instructors attain full access to the "Repository of 3D Objects", upon which, all of the Experiments of the Library are constituted. They are, therefore, capable of composing new experiments, for their learners, from scratch, without the necessity of advanced computational skills, to reason why *howlearn* also functions as an "Authoring Gamification Tool".

Last, but not least, the System is in complete accordance with the WCAG 2.1 Web Content Accessibility Guidelines, making *howlearn* accessible to a wider range of people with disabilities, such as color blindness, photosensitivity, hearing loss and cognitive limitations (color palette of the entire user experience: AAA—"Excellent"—colors: #FFC845 #1B365D #007478 #FFFFF #00000—Raftopoulou & Pallis, 2023b).

The innovation, however, within *howlearn* lies in the "Personalized Feedback Report Dashboard" that all of its users receive, upon completion of their Experiments. More particularly, Machine Learning Algorithms, in union with more traditional Statistical Analysis Algorithms, constitute the backbone of the personalized feedback process. These algorithms, the selection of which, is based on their performance and suitability/applicability, within *howlearn*, include, but are not limited to, various unsupervised Clustering Algorithms (such as K-Means), which are utilized, in order to discover hidden patterns within the learner population, with the intention of grouping together, not only students of similar learning capabilities or interests, but also Experiments that share common distinctive traits.

The results occurring from these Clustering Algorithms are being, further, improved and extended, when used in combination with Statistical Analysis Algorithms, in order to accurately determine the dynamic learner subset each student belongs to, and map out the correlation between groups of learners and their learning performance. Ultimately, learners receive 3 distinctive Recommendations (namely, 3 distinctive Recommendation Systems), on future Experiments they should fulfil, based on:

- what other learners, similar to them (in terms of their Experiment Preferences) like (Collaborative Filtering Recommendation System);
- their Experiment Performance (Collaborative Filtering Recommendation System);
- other experiments they might enjoy, according to their own Experiment Preferences (Content-Based Filtering Recommendation System).

Literature Review: Recommendation Systems

Recommendation Systems Overview

A Recommendation System refers to a subset of information filtering system, supplying recommendations for items which are likely to be relevant and suitable to a particular user (Ricci, Rokach, & Shapira, 2022; Grossman, 2010). Commonly, these recommendations touch upon decision-making processes, which may, for instance, refer to a potential product of purchase or a probable music piece to listen to (Ricci et al., 2022). Hence, such systems are immensely useful, when a user is requested to choose a particular item, from a feasibly overwhelming number of potential items, offered by the system (Ricci et al., 2022; Resnick & Varian, 1997).

It, therefore, becomes elucidated that, Recommendation Systems may be applied in a multitude of fields, with some of their most well-known applications revolving around the music industry (automated playlist generators), or the social media/web-based content industry (Gupta et al., 2013; Baran, Dziech, & Zeja, 2018). These systems are capable of operating using just a unique input (i.e. music), or numerous inputs, between and within web networks (i.e. articles and search queries). However, their application is also fully functional when it comes to exploring research articles (Chen, Ororbia II, & Giles, 2015), colleagues (Chen, Gou, Zhang, & Giles, 2011) or, even, financial services (Felfernig, Isak, Szabo, & Zachar, 2007).

Recommendation Systems employ either Collaborative Filtering and Content-Based Filtering, or even both (Jafarkarimi, Sim, & Saadatdoost, 2012). Collaborative Filtering proceeds with the construction of a model, based on users' previous behaviors (items preferred in the past or scores/ratings provided to them), as well as based on equivalent choices, made by other users. The model in question is, consequently, utilized to produce item predictions (or item ratings), which would be of particular interest to the user (Melville & Sindhwani, 2010). Content-Based Filtering, on the other hand, makes use of discrete, pre-defined item characteristics, in order for subsequent items with related properties to be recommended (Mooney & Roy, 1999).

Collaborative Filtering

Collaborative Filtering (CF) is a Recommendation Systems approach (Ricci et al., 2022), mostly used with the intention to make automated forecasts (filtering) on an individual's interests, by assessing preferences of numerous other individuals (collaborating). The fundamental consideration of this technique lies within the following statement: if two people, A and B, share the same opinion, on an issue of interest, X, then A would be much more likely to share B's opinion upon another matter of interest, say Z, in comparison to the opinion of a third individual, say C, randomly chosen. Hence, such a Recommendation System, for preferences on an item of interest X, produces predictions about items of interest an individual would, most probably, enjoy, given a limited list of their very own preferences (likes/dislikes/item rankings—Resnick & Varian, 1997). Although these forecasts are user-specific, they utilize information derived from numerous individuals.

Content-Based Filtering

Content-Based Filtering is an alternative Recommendation Systems technique, making use of an illustration of the specific item in question and a profile describing the user's preferences (Aggarwal, 2016; Brusilovsky, Kobsa, & Nejdl, 2007). This filtering method is preferred when there is enough data on the main element of interest, yet, not on the user. In other words, the generated recommendation is treated as a user-centric classification approach, in which a classifier learns for the end user's preferences, by taking into account an element's attributes and characteristics.

Therefore, such systems function based on keywords, usually applied as short descriptions to the elements in question, as well as on a well-defined user profile, which is formulated with the intention to demonstrate the kind of elements that this user prefers. In a nutshell, Content-Based Filtering attempts to suggest recommendations of elements/items, similar to the ones that the person in question formerly enjoyed or is enjoying at this moment in time. More specifically, numerous contestant elements are being contrasted with elements rated in the past by the user, process upon which, the best-matching elements are the ones being recommended. To conclude, the user's profile is primarily extracted based on inputs concerning the user's likes and dislikes, as well as on a brief history of the interactivity of the latter, with the Recommendation System itself.

Personalized Feedback Report Dashboard—A Case Study: *howlearn*

Data Analytics of In-Game Metrics—Dashboard of Personalized User Performance

In order for *howlearn* to be able to provide its users with a highly informative Personalized Feedback Report Performance Dashboard, the storage of In-Game Player Metrics was deemed necessary. These metrics summarize learners' performance, by providing detailed information on the in-game Experiment competency of the latter, and on whether they have enjoyed that very Experiment or not, as well as on the kind of Experiment currently fulfilled by the learner in question (for more information, see Table 1 below):

Table 1

Game Metrics Stored Within howlearn's Database

Variable	Values
PlayerID	Player Identifier
ExperimentID	32 Experiment Names/Identifiers
ExperimentCategory	ICT STEM Social Sciences, Humanities and Art Entrepreneurship and Innovation Health and Sports Foreign Languages
IsOpenSpaceExperiment	TRUE (Movement within the Space) FALSE (Static Camera)
IsCharacterReactiveExperiment	TRUE (Avatar-Based Interaction) FALSE (Non-Avatar-Based Interaction)
UserExperienceRating	Rating: 1 to 5 Stars
ExperimentTimeRunning	Total Experiment Time (Duration of Experiment)
CompletionPercentage	Percentage of Experiment Completed
Success	TRUE (Experiment Successfully Completed) FALSE (Failure of Experiment Completion)
Answers	TRUE (Correct Answer to the equivalent Experiment Question) FALSE (Wrong Answer to the equivalent Experiment Question)
Date	Date and Time of Experiment Conduction

Note. Source: Authors (2023).

The main idea behind the pre-mentioned metrics lies within the system's effort to provide the Gamified Education Community with metrics that are Universal to the world of Gamification in Education (i.e. "ExperimentCategory", "IsOpenSpaceExperiment"), so that the Dashboard of Personalized User Performance and its Recommendation Systems may be easily Reproducible (hence providing a Solution to the

“Reproducibility Issue”, from which Recommendation Systems suffer to this day—for more information, see Discussion—Reproducibility below).

In terms of the In-Game Data Analytics Personalized Feedback Report Dashboard received by all learners upon completion of their Experiments, it includes detailed Data Visualizations of (for more information, see Figure 1 below):

- the comparison between their own time-related performance and the average time-related performance of all students (on the Experiment in question);
- the Experiment completion percentage;
- the amount of correct and incorrect answers to all of the questions of the gamified scenario in question;
- the, relative to all of the other learners, learner’s score, represented as the kth Percentile (meaning that, this very learner scores better/greater than K% of learners who have taken up that particular Experiment of interest). Percentiles, here, inform us on how a specific learner of interest compares to other learners (Frost, 2019).

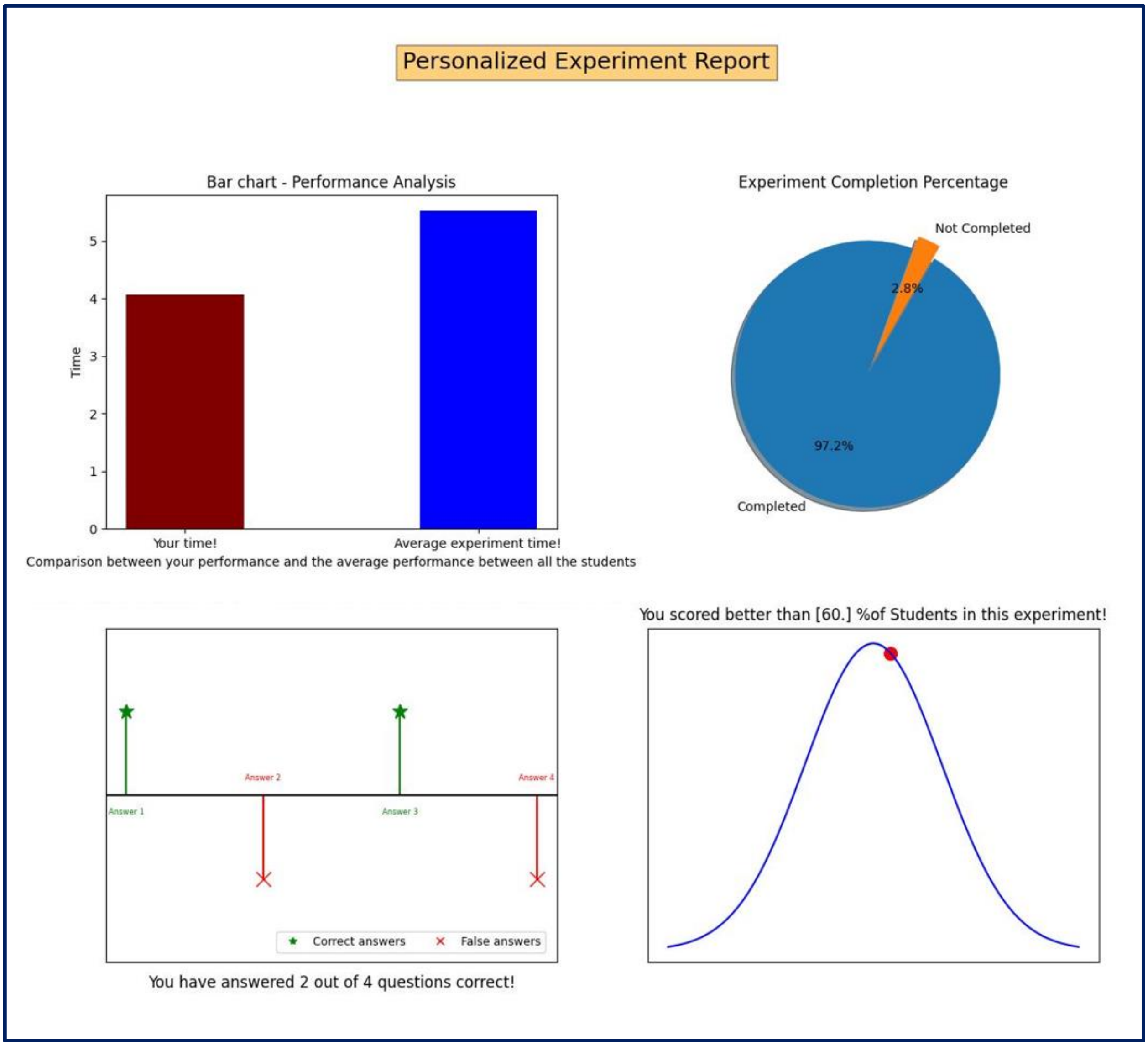

Figure 1. Learner’s in-game data analytics personalized feedback report dashboard. Source: Authors (2023).

Formulating an Innovative Triple Recommendation System, for the Societal Innovative Gamified Learning System, *howlearn*

First Custom Recommendation System—Collaborative Filtering Recommendation System Using Cosine Similarities and K-Means Clustering Algorithms

The very first Custom Recommendation System of *howlearn* is based on learners' User Experience Rating of the Experiment they have just fulfilled. More particularly, upon a table of 3 Columns (variables: "PlayerID", "ExperimentID" and "UserExperienceRating"), the System's algorithms look for cosine similarities, that is, mathematical representations providing a useful measure of how similar two documents (learners in our case) are likely to be, in terms of their subject matter (rating of Experiments fulfilled), and independently of the length of the documents (Singhal, 2001).

A K-Means Clustering Algorithm is then being applied (vector quantization method), aiming to partition our n observations (learners) into k Clusters (5, in our case), in which, each observation belongs to the Cluster with the nearest mean (Cluster Centroid), serving as a precursor of the Cluster (Forgy, 1965).

This way, supposing we have a main "observation learner of interest", they would be clustered together with those learners, who appear to have a similar Experiment Rating behavior to them. In other words, the System manages to figure out the "Experiment Rating Cluster" in which our "observation learner" belongs to and places them within it.

Once this process is finalized, an average rating is provided to all of the Experiments fulfilled by learners belonging to the same Cluster. Based on that average rating and by taking into consideration the Experiments, within each Cluster, which have been fulfilled by the other learners of that Cluster, yet not by our "observation learner" in question, we extract Experiments, as Recommendations to the latter, for future realization.

It, hence becomes clear that, the first Custom Recommendation System of *howlearn* is a Collaborative Filtering Recommendation System, since, our player ("observation learner of interest") receives recommendations of Experiments that they would most likely enjoy, based on the fact that, these Experiments have been highly rated (and, hence, enjoyed) by other players, who share a similar Experiment rating behavior with the "observation learner" in question. The Experiment Recommendations are, finally, being sorted, and the top matches (Experiments with the highest of ratings) are being suggested to the "observation learner", upon exclusion of the Experiments that they have already completed (Recommendation Based on What Other Learners Like You Liked—for more information, see Figure 2 below).

Second Custom Recommendation System—Collaborative Filtering Recommendation System Using Statistical Analysis (Standard Scores—Z-Scores)

The standard score (otherwise known as the z-score) refers to the number of standard deviations by which the value of a raw score is above or below the mean value of what is being observed or measured. It is calculated by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation (Kreyszig, 1979). In our case, since the population of *howlearn*'s end users is unknown, we are using the sample mean and sample standard deviation, as estimates of the population mean and population standard deviation for our variables of interest.

The second Custom Recommendation System of *howlearn* is, once again, a Collaborative Filtering Recommendation System, although, this time, it makes use of a Statistical Analysis approach, based on the extraction of standard scores (z-scores) related to the "ExperimentTimeRunning" and "CompletionPercentage"

In-Game Metrics (the latter of which, also takes into consideration the "Answers" and "Success" In-Game Metrics).

More particularly, the time (throughout which an Experiment runs) is being scaled (transformation of our data on a 0-1 scale), process upon which, we formulate the z-scores for our "ExperimentTimeRunning" and "CompletionPercentage" variables of interest (for more information, see Equation 1, Equation 2, Equation 3 and Equation 4 below). This way, our System manages to identify/tell apart, high-performing learners from learners who are performing poorly, by assessing how distant they are, from the mean of the distribution in question (we assume that Learners' Performance and Experiments' Performance, both follow a Normal Distribution). The extracted z-scores, thence, result from the following calculations (n represents *howlearn*'s learners' sample size):

$$z_{score_{ExperimentTimeRunning}} = \frac{ExperimentTimeRunning - mean\ \ (ExperimentTimeRunning)}{stdev(ExperimentTimeRunning)} \quad (1)$$

where

$$stdev = \sqrt{\frac{\sum_{i=1}^{n}(ExperimentTimeRunning_i - mean(ExperimentTimeRunning))^2}{n-1}} \quad (2)$$

and $ExperimentTimeRunning_i$ is the ith Experiment Time Running observation.

$$z_{score_{CompletionPercentage}} = \frac{CompletionPercentage - mean\ \ (CompletionPercentage)}{stdev(CompletionPercentage)} \quad (3)$$

where

$$stdev = \sqrt{\frac{\sum_{i=1}^{n}(CompletionPercentage_i - mean(CompletionPercentage))^2}{n-1}} \quad (4)$$

and $CompletionPercentage_i$ is the ith Completion Percentage observation.

We, then, proceed with the calculation of a Weighted Sum of the z-scores:

$$Weighted\ \ Score = \frac{30}{100} * z_{score_{ExperimentTimeRunning}} + \frac{70}{100} * z_{score_{CompletionPercentage}} \quad (5)$$

so that learners taking their time, in order to properly finalize their Experiments, are not being "punished" by the System (for more information, see Equation 5 above).

Once we scale our extracted values, we proceed with the calculation of the mean score of every Experiment, as well as the calculation of the mean score of every Player, process, ultimately allowing us to rank learners, based on the "Learners' Dynamics", whilst, also ranking Experiments based on the "Experiments' Dynamics". In other words, we end up with 2 Custom Statistical Metrics, the matching of which, is of ultimate interest to us.

More precisely, initially, the mean score of learners (PlayerID) is defined and learners (players) are classified in comparison to the rest of the learners, according to the Percentile Rank of their received score. Similarly, the mean score of the Experiments is being retrieved. Of course, since high values of Experiments' z-scores imply that the Experiments in question are easier to be fulfilled (since they acquire higher completion percentages), we had to reverse these scores, so that high-performing learners receive recommendations of more demanding (harder to be successfully completed) Experiments and so that, poorly performing learners, are not overwhelmed with Experiment Recommendations, which, by far, exceed their potential (such recommendations would, massively, discourage learners from the act of active participation in the learning process).

Consequently, learners who are positioned in the middle (mean) of the Normal Distribution of Learners' Performance receive recommendations of Experiments which are also positioned in the middle of the Normal Distribution of Experiments' Performance. In other words, learners are being recommended to fulfil those Experiments that are the closest, in distance, to the area within which they are positioned, on their Normal Distribution, which is how Learners' Dynamics, ultimately, sensibly match Experiments' (Modules) Dynamics (Recommendation Based on Your Performance—for more information, see Figure 2 below).

As a matter of course, Experiments already completed by the learner in question are not taken into consideration and are, therefore, no longer available as recommendations, thanks to the System's cross-checking capability, between Experiment realizations and Experiment suggestions, which ensures that only Experiments never taken by the learner in question, would ever be recommended to them (unique Experiment suggestions feature of *howlearn*).

Third Custom Recommendation System—Content-Based Filtering Recommendation System

The third Custom Recommendation System of *howlearn* is a Content-Based Filtering Recommendation System. More precisely, the System takes note of the Experiments already completed by the learner in question, as well as of the rating that the latter has provided them with. The Experiments are, then, being classified, based on their unique identifying qualities (i.e. "ExperimentCategory", "IsOpenSpaceExperiment", "IsCharacterReactiveExperiment"), once again, by making use of cosine similarities, so as to define "similar, in attributes/characteristics, Experiments".

The System, then, examines which Experiments have received the highest of ratings by the learner and, by, additionally, assessing their equivalent cosine similarities, it goes ahead with recommending, for future realization, the highest-rated ones (which have not yet been completed/taken up by the learner in question) and which, additionally, share distinctive traits of high interest to the learner (since that specific learner seems to enjoy fulfilling Experiments portraying those particular traits/attributes). This approach, thence constitutes a Content-Based Filtering Custom Recommendation System, since, this time, it is similar Experiments (Content) that are taken into account, not similar learners (Collaborative) (Recommendation Based on Other Experiments You Might Enjoy, According to Your Preferences—for more information, see Figure 2 below).

Conclusions

Social Impact of Formulating Custom Recommendation Systems on Innovative 3D Gamified Learning Systems

The integrated solution proposed in the framework of the "Innovative Research Project *howlearn*" provides significant social benefits, since it allows for the effective enforcement of sought-after technological advancements, to the whole of the educational community, in the form of a cross-platform learning system, capable of operating, both online and offline, therefore overcoming potential geographical and/or other socio-economic educational boundaries, hence allowing all learners to attain access to equal, high-end, innovative education.

Moreover, *howlearn*'s Custom Recommendation Systems give learners the opportunity to navigate themselves through a wide range of educational material, in a concise, clairvoyant manner, through suggestions for future Experiment realizations, which holistically enhance their knowledge and educational potential; not only are these recommendations in accordance with learners' preferences or weaknesses (general performance), they, additionally, account for general class competency and class perception (rating) of the Experiments in question, whilst also assessing the distinctive traits/attributes of each and every single one in the System's Experiments.

As a result, learners get to ameliorate their educational performance (enhancement of hard skills), while also gaining a better understanding of how they perform, in comparison to the rest of the class, or, of how, classmates who share similar preferences with them, perceive other educational material, which would, most likely, be of interest to the former too (enhancement of soft skills, i.e. how learners work and interact with the rest of the class—for more information, see Figure 2 below). Needless to say, learners commence to acquire a better understanding of their very own data (Data Analysis Skills), therefore becoming data literate and digitally upskilled.

Consequently, by gradually building on their technical acumen and grit, while remaining engaged in the learning process (constant curiosity over what educational topic the next recommendation could introduce them to), learners end up receiving state-of-the-art, multidimensional (cognitively, emotionally, socially and technologically), sustainable, inclusive (accessibility friendly) education.

Figure 2. Custom collaborative filtering and content-based filtering recommendation systems of *howlearn*. Source: Authors (2023).

Discussion

Reproducibility

Recommender systems are rather infamous for not being able to assess offline, which leads to a "Recommendation Systems-related, publications reproducibility crisis". More particularly, a 2019 paper surveyed a small-scale of state-of-the-art publications employing deep learning or neural methods to the top-k recommendation problem (i.e. SIGIR, IJCAI), showing that, on average, less than 40% of the publications could be reproducible, by the authors of the survey. The paper examines a handful of probable research-related issues and proceeds with the suggestion of ameliorated scientific practices in those areas of interest (Ferrari Dacrema, Boglio, Cremonesi, & Jannach, 2021; Ferrari Dacrema, Cremonesi, & Jannach, 2019; Rendle, Krichene, Zhang, & Anderson, 2020).

Some of the latest findings on similar methods' benchmarking, lead to qualitatively vastly differing results (Sun et al., 2020), whereas neural methods were considered to be among the state-of-the-art suggested methods. This result came as no surprise, given the fact that, together with deep learning for recommender systems, neural methods are massively being applied in numerous contemporary recommender system challenges, WSDM (Schifferer et al., 2021), RecSys Challenge (Volkovs et al., 2018), especially in the industry areas, where they are being exhaustively tested (Raimond & Basilico, 2018; Chen et al., 2015; Ma, Narayanaswamy, Lin, & Ding, 2020).

Nevertheless, it is worth mentioning that reproducibility in Recommendation Systems has been an issue for a while. In 2011, it was censured that "the extensions and reproducibility of Recommendation Systems' research results are rather strenuous, with evaluations not being assessed steadily" (Ekstrand, Ludwig, Konstan, & Riedl, 2011). Two other researchers inferred that

> research associated with Recommendation Systems is suffering from a crisis, since, the results presented in numerous of those Papers contribute little knowledge to the community, mainly since their results cannot easily be assessed, therefore failing to supply insightful benefactions to the community. (Konstan & Adomavicius, 2013)

Consequently, a lot of the research conducted in the area of Recommendation Systems may as well be contemplated as non-reproducible (Beel, Breitinger, Langer, Lommatzsch, & Gipp, 2016).

As a result, it is often difficult to have a clear indicator of which recommendation technique to apply in a specific Recommendation Systems problem. Said and Bellogín (2014), benchmarked some of the most sought-after frameworks for recommendations, only to end up with massive inconsistencies in the results, even in cases where the exact same algorithmic approaches and datasets were being utilized. A few other researchers illustrated how, even the slightest of variations in the Recommendation Systems Algorithms, were capable of leading to the vast decrease of the overall effectiveness of the recommender system (Beel et al., 2016). They, henceforth, presumed that seven actions are deemed necessary, for the situation to be ameliorated: (1) survey other research fields, so as to absorb knowledge from them, (2) uniformly fathom reproducibility, (3) recognize and gain understanding of the principal components affecting reproducibility, (4) regulate more extensive experimentations, (5) streamline publication practices, (6) foster the evolvement and usage of recommendation frameworks, and (7) inaugurate best-practice guidelines for Recommendation Systems research.

Taking into consideration the difficulty in creating Reproducible Recommendation Systems and having gained great understanding of the notions of Gamification and Game Metrics (In-Game Data Analytics) in education, as part of the process of constituting *howlearn*, we have proceeded (as mentioned in Personalized Feedback Report Dashboard—A Case Study: *howlearn*—Data Analytics of In-Game Metrics—Dashboard of Personalized User Performance above) with the formulation of Universal In-Game Metrics (In-Game Data Analytics) for *howlearn*, so that our Custom Recommendation Systems provide results, easily and holistically reproducible, within any educational gamified context.

In fact, we suggest that, all of the Game Metrics and Custom Recommendation Systems of *howlearn*, presented in this paper, may be used as a set of pre-defined, Optimal Universal In-Game Metrics suggestions, leading to highly Reproducible Recommendation Systems, accompanying any Innovative Gamified Learning System like *howlearn*.

Hybrid Recommendation Systems

Although Collaborative Filtering and Content-Based Filtering are two of the predominant Recommendation Systems techniques, nowadays, a great number of Recommendation Systems apply the so-called Hybrid technique, leading to "Hybrid Recommendation Systems", amalgamating Collaborative Filtering and Content-Based Filtering.

Hybrid Filtering may be executed in various different ways: by applying content-based and collaborative-based forecasts one by one and merging them afterwards; by inserting content-based competencies to a collaborative-based technique (and vice versa); or by coalescing the techniques into one, unified model (Adomavicius & Tuzhilin, 2005).

It is worth mentioning that, studies empirically comparing the performance of Hybrid Recommendation Systems with pure Collaborative Filtering or Content-Based Filtering Recommendation Systems have exhibited the outperformance of the latter by hybrid approaches, with regard to the overall Accuracy of the results produced by the models. Additionally, the application of hybrid techniques may also provide solutions with regard to some of Recommendation Systems' major issues (i.e. their "cold start issue" or their "sparsity issue"—Hoekstra, 2010).

Netflix is highly applying Hybrid Recommendation Systems, since the website's recommendations are the result of the comparison between the watching and searching habits of similar individuals (Collaborative Filtering), as well as the suggestion of movies that possess similar attributes with movies that the individual in question has already provided high ratings with (Content-Based Filtering—Gomez-Uribe & Hunt, 2015).

Consequently, the application of Hybrid Techniques could constitute a beneficial future implementation within *howlearn*, further improving the Accuracy of the produced Recommendations, especially should Feature Combination is applied. In this technique, features derived from different knowledge sources would be combined together and given to a single recommendation algorithm (Zamanzadeh Darban & Valipour, 2022). If, concurrently, these different knowledge sources are the result of *howlearn*'s Official Pilot Testing Release on its end users, then, an extensive Data Analysis on the perceptions of the latter, with respect to the System, could feed our algorithms with even more variables of main interest (Principal Components), hence, leading the way towards our System's Optimization, thence comprising *howlearn*'s upcoming milestone.

References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering, 17*(6), 734-749.

Aggarwal, C. C. (2016). *Recommender systems: The textbook*. Springer.

Baran, R., Dziech, A., & Zeja, A. (2018). A capable multimedia content discovery platform based on visual content analysis and intelligent data enrichment. *Multimedia Tools and Applications, 77*(11), 14077-14091.

Beel, J., Breitinger, C., Langer, S., Lommatzsch, A., & Gipp, B. (2016). Towards reproducibility in recommender-systems research. *User Modeling and User-Adapted Interaction, 26*(1), 69-101.

Brusilovsky, P., Kobsa, A., & Nejdl, W. (2007). *The adaptive web: Methods and strategies of web personalization*. Springer.

Chen, H.-H., Ororbia II, A. G., & Giles, C. L. (2015). *ExpertSeer: A keyphrase based expert recommender for digital libraries*. Retrieved from https://arxiv.org/abs/1511.02058

Chen, H.-H., Gou, L., Zhang, X., & Giles, C. L. (2011). CollabSeer: A search engine for collaboration discovery. In *JCDL'11: Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries* (pp. 231-240).

Ekstrand, M. D., Ludwig, M., Konstan, J. A., & Riedl, J. T. (2011). Rethinking the recommender research ecosystem: Reproducibility, openness, and LensKit. In *RecSys'11: Proceedings of the Fifth ACM Conference on Recommender Systems* (pp. 133-140).

Felfernig, A., Isak, K., Szabo, K., & Zachar, P. (2007). The VITA financial services sales support environment. In *IAAI'07: Proceedings of the 19th National Conference on Innovative Applications of Artificial Intelligence* (pp. 1692-1699).

Ferrari Dacrema, M., Boglio, S., Cremonesi, P., & Jannach, D. (2021). A troubling analysis of reproducibility and progress in recommender systems research. *ACM Transactions on Information Systems, 39*(2), 1-49.

Ferrari Dacrema, M., Cremonesi, P., & Jannach, D. (2019). Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *RecSys'19: Proceedings of the 13th ACM Conference on Recommender Systems* (pp. 101-109).

Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics, 21*(3), 768-769.

Frost, J. (2019). *Percentiles: Interpretations and calculations*. Retrieved from https://statisticsbyjim.com/basics/percentiles/

Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems, 6*(4), 1-19.

Grossman, L. (2010). *How computers know what we want—Before we do*. Retrieved from https://web.archive.org/web/20100530064045/http:/www.time.com/time/magazine/article/0,9171,1992403,00.html

Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D., & Zadeh, R. B. (2013). WTF: The who to follow service at Twitter. In *WWW'13: Proceedings of the 22nd International Conference on World Wide Web* (pp. 505-514).

Hoekstra, R. (2010). *The knowledge reengineering bottleneck*. Semantic Web – Interoperability, Usability, Applicability 1 (2010) 1, IOS Press. Retrieved from https://www.semantic-web-journal.net/sites/default/files/swj32.pdf

Jafarkarimi, H., Sim, A. T. H., & Saadatdoost, R. (2012). A naïve recommendation model for large databases. *International Journal of Information and Education Technology, 2*(3), 216-219.

Konstan, J. A., & Adomavicius, G. (2013). Toward identification and adoption of best practices in algorithmic recommender systems research. In *RepSys'13: Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation* (pp. 23-28).

Kreyszig, E. (1979). *Advanced engineering mathematics* (4th ed.). Wiley.

Ma, Y., Narayanaswamy, B., Lin, H., & Ding, H. (2020). Temporal-contextual recommendation in real-time. In *KDD'20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2291-2299).

Melville, P., & Sindhwani, V. (2010). *Recommender systems*. Encyclopedia of Machine Learning. Retrieved from https://www.prem-melville.com/publications/recommender-systems-eml2010.pdf

Mooney, R. J., & Roy, L. (1999). Content-based book recommending using learning for text categorization. Proceedings from the *SIGIR-99 Workshop on Recommender Systems: Algorithms and Evaluation*. Berkeley, CA, August 1999. Retrieved from https://www.cs.utexas.edu/users/ml/papers/libra-sigir-wkshp-99.pdf

Raftopoulou, N.-M., & Pallis, P. L. (2023a). Formulating innovative societal educational learning systems using gamification techniques in 3D environments. *Sociology Study, 13*(1), 25-37.

Raftopoulou, N.-M., & Pallis, P. L. (2023b). Establishing an appealing cross-platform innovative educational gamified learning system product: A complete user experience case study. Proceedings from the *4th Global Conference on Education and Teaching*. Athens, Greece.

Raimond, Y., & Basilico, J. (2018). *Deep learning for recommender systems*. Deep Learning Re-Work SF Summit. Retrieved from https://www.slideshare.net/moustaki/deep-learning-for-recommender-systems-86752234

Rendle, S., Krichene, W., Zhang, L., & Anderson, J. (2020). Neural collaborative filtering vs. matrix factorization revisited. In *RecSys'20: Proceedings of the 14th ACM Conference on Recommender Systems* (pp. 240-248).

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM, 40*(3), 56-58.

Ricci, F., Rokach, L., & Shapira, B. (2022). Recommender systems: Techniques, applications, and challenges. In F. Ricci, L. Rokach, and B. Shapira (Eds.), *Recommender systems handbook* (3rd ed., pp. 1-35). New York: Springer.

Said, A., & Bellogín, A. (2014). Comparative recommender system evaluation: Benchmarking recommendation frameworks. In *RecSys'14: Proceedings of the 8th ACM Conference on Recommender Systems* (pp. 129-136).

Schifferer, B., Deotte, C., Puget, J.-F., de Souza Pereira, G., Titericz, G., Liu, J., & Ak, R. (2021). Using deep learning to win the Booking.com WSDM WebTour21 challenge on sequential recommendations. Proceedings from the *WSDM'21: ACM Conference on Web Search and Data Mining WebTour*. March 12, Jerusalem, Israel.

Singhal, A. (2001). Modern information retrieval: A brief overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 24*(4), 35-43.

Sun, Z., Yu, D., Fang, H., Yang, J., Qu, X., Zhang, J., & Geng, C. (2020). Are we evaluating rigorously? Benchmarking recommendation for reproducible evaluation and fair comparison. In *RecSys'20: Proceedings of the 14th ACM Conference on Recommender Systems* (pp. 23-32).

Volkovs, M., Rai, H., Cheng, Z., Wu, G., Lu, Y., & Sanner, S. (2018). Two-stage model for automatic playlist continuation at scale. In *RecSys Challenge'18: Proceedings of the ACM Recommender Systems Challenge 2018* (pp. 1-6).

Zamanzadeh Darban, Z., & Valipour, M. H. (2022). GHRS: Graph-based hybrid recommendation system with application to movie recommendation. *Expert Systems with Applications, 200*, 116850.