Let the Games Speak by Themselves: Towards Game Features Discovery Through Data-Driven Analysis and Explainable AI

Riccardo Martoglia[1,2], Matteo Pontiroli[1]
[1] *FIM - University of Modena and Reggio Emilia, Italy*
[2] *GAME Science Research Center, Italy*
riccardo.martoglia@unimore.it, mpontiroli83@gmail.com

***Abstract*—The idea behind this work is to start exploring the application of data analytics and (explainable) machine learning techniques to better understand games and discover new features that will possibly help in effectively exploiting them in different socially useful domains. We prove the feasibility of the idea by: (i) collecting a large dataset of board game information; (ii) designing and testing an information processing pipeline for automatically discovering game categories and game mechanics, with some first encouraging results. In the future, we plan to further generalize this approach for different kinds of games and for discovering currently unknown but useful aspects, e.g. games or game features that could better foster Computational Thinking in education, those better suited to be applied in social distancing contexts, and so on.**

***Index Terms*—Game analytics, (explainable) machine learning, text analysis, automatic categorization**

I. Introduction

Games are being applied more and more frequently to an ever growing number of domains in our everyday lives, from entertainment to education, from creativity to technology. Indeed, game-related research is in these days a very hot field of study, in particular for many areas related to computer science including data analytics and artificial intelligence studies for computer-assisted game design [3] or for teaching computers play games [11]. Nonetheless, the enormous potential of games is still largely unutilized and it feels there is much left to explore about the benefits of game-related applications, especially in non strictly entertainment-related contexts.

In this paper we present the first steps of a work-in-progress research we are conducting at Game Science Research Center[1], an interuniversity research center aimed at promoting, supporting and spreading the research in the field of Game Science. The aim: employing data analytics and machine learning techniques to better understand games and ultimately discover new features that can help in effectively exploiting them in different socially useful domains. In this work we prove the feasibility of the idea by specifically focusing on board games and on some preliminary objectives: a) we collect a large dataset of 50000 board games by extracting relevant data from the reference site in this field, BoardGameGeek[2] (BGG); b) we perform a series of experiments where we test how effectively different machine learning techniques can help in automatically discovering game categories and game mechanics using only a short textual description of the game (the one that is usually printed on the back of the box) and show some first encouraging results. As done in the past in different analytics scenarios [2], [4], we also cosider the application of *explainable machine learning* techniques so to "explain" the results in meaningful ways.

While game categories and mechanics are currently available on the BGG game records, their automatic discovery could certainly speed up the experts' work in manually categorizing them. But most of all, we plan to generalize this approach in future research and apply it to additional inputs given by game experts, expanding the scale of their intuitions to effectively discover currently unknown aspects of games: for instance, games or game features that could better foster Computational Thinking in education, or those better suited to be applied in social distancing contexts. In this way, we hope that this *data-driven* approach will hopefully facilitate the generalization of ”rules” or ”findings” to be further inspected and validated by experts and that would not have been possible (or foreseeable) otherwise.

The rest of the paper is organized as follows: in Section II we briefly report on related work; Section III is devoted to data collection, while Section IV and V describe the proposed approach and detail the obtained results, respectively. Finally, conclusions and future work are presented in Section VI.

II. Related Work

There is a very large body of work on the use of machine learning and data analytics techniques on game related scenarios, an area that is becoming increasingly popular especially in recent times. The aims are very much disparate, from computer-assisted game design [3] to AI powered game playing [10], [11], with important applications also on non strictly entertainment-related contexts. More specifically, the analysis of board game information has been an important but still not widely explored area, with some notable researches that have been performed both in academic and non-academic contexts.

Speaking about board game information, Board Game Geek is certainly acknowledged as one of the primary information

[1]https://gamescience.imtlucca.it
[2]http://boardgamegeek.com/

sources and many data analysts and researchers worked on its data for various objectives. Most of the efforts have been made to create recommendation systems, a fundamental component also present in many e-commerce sites such as Amazon, and Ebay. An example of an academic research for a recommendation system exploiting BGG data is [16], while a notable one in a non-academic context is Board Game Finder, available as a website [14]. The authors focused on collaborative filtering and used a machine learning approach to give estimations about users' unrated games and consequently suggest them new games. The employed attributes from BGG are the minimum and maximum number of players, the time taken for a game, the designers of a game (which proved to be the most important predictor of the like score), the category and mechanics of the game. Other works on board game analytics and exploiting BGG data are focused on different objectives, from the prediction of board games review ratings [9] to the creation of an ontology of boardgame mechanics based on the MDA Framework [12], to the definition of archetypes of players and games and their relation to game mechanics and genres [15].

The research we present in this paper is certainly in some ways related to the above discussed ones; however, our final aim is different and quite novel, since we want to investigate the feasibility of applying data analytics and machine learning techniques for automatically discovering game features, starting only from a short textual description. Besides the presented analyses, a further contribution of this work is the dataset: being this a first step toward future generalization of the approach to discover unknown aspects of games, also exploiting further inputs and targets, the dataset we collect from BGG for 50000 boardgames goes beyond the games' textual descriptions and contains a larger number of data / fields than the ones usually found in similar research (see Section III for a short description).

We complete this related work analysis with some final observations on one of the aspects that characterizes and distinguishes the research proposed in this paper, i.e., the use of interpretable machine learning techniques. In general, there has been a recent emphasis on the need for explanations of machine learning systems [7] in many scenarios, in particular in the medical setting, where clinicians need to know the "reasons" behind ML-based predictions (the specialized research trend is Interpretable Machine Learning in Healthcare [1]). One important aim in this research is to perform the first steps in going beyond the black box nature of machine learning predictions also in the game analytics context, where this has seldom been done.

III. Data Collection and Dataset Creation

For the purpose of this research (and its future developments) we created two datasets by extracting the required information from the BGG website. The datasets have different sizes but share the same structure, and include detailed information about the 50000 (large dataset, 5000 for the small dataset) top-rated board games in the website top board game ranking. The small dataset is a subset of the large one; the idea behind the use of two datasets is to study the impact of the number of samples over the models' prediction capability.

	name	description	boardgamemechanic	boardgamecategory
23160	Speed and Steel	Speed and steel is a two player game which sim...	Dice Rolling,Hexagon Grid	Territory Building,Wargame,World War II
38233	Scouting	This game with a copyright date of 1926 is cal...	Roll / Spin and Move	Adventure,Children's Game
42919	Veto! CCG	Veto! is a Polish CCG game. Players lead they ...	Card Drafting,Hand Management	Card Game,Collectible Components

TABLE I
Samples from the dataset showing game name, description, mechanics and categories

The datasets were generated by using an ad-hoc python script that downloads the webpages and parses the data contained in them, producing a standard CSV format for compatibility. The scraped data include the following information: *name* (english name of the boardgame), *description* (short textual description as found on the game box), *boardgamemechanic* (list of game mechanics present in the game, such as "dice rolling" or "card drafting"), *boardgamecategory* (list of game categories, such as "wargame" or "adventure"), *boardgamefamily* (list of families of games which the game belongs to), *type* (to discriminate board games from expansions), *yearpublished* (publishing year), *minplayer* and *maxplayer* (minimum and maximum number of players), *playingtime*, *minplaytime* and *maxplaytime* (typical, minimum and maximum playtime, respectively), *minage* (minimum age for playing), *boardgamedesigner*, *boardgameartist* and *boardgamepublisher* (game authors/publishers), *boardgameexpansion* (available expansions), *languagedependence* (a number in range [1-5] indicating the importance of written text for the game playability). In particular, while the dataset rich schema is devised so to support a number of future analysis, the tests presented in this paper will be focused on the first four fields: see Table I for a small sample. Notice that a game can have more than one mechanic, the same is for the categories.

IV. Data Processing, Analysis and Models

The main goal of the analyses we present in this paper is to test the feasibility of an automatic classification approach capable of assigning the correct game mechanics/categories using only the game description as input feature, measuring the performances obtained for various configurations and models. Figure 1 provides an overview of the proposed approach. The details will be discussed in the following of this section; after a short exploration of the dataset, we will describe its processing and encoding (top part of figure), then we will present the actual testing phase (lower part of figure) and discuss the obtained results. The tools employed for our analysis include *Pandas*[3] for data manipulation, *NLTK*[4] for natural language processing and *scikit-learn*[5] for feature extraction, dataset

[3]Pandas, http://pandas.pydata.org/
[4]Natuaral Language Toolkit, https://www.nltk.org/
[5]Scikit-learn, http://scikit-learn.org/

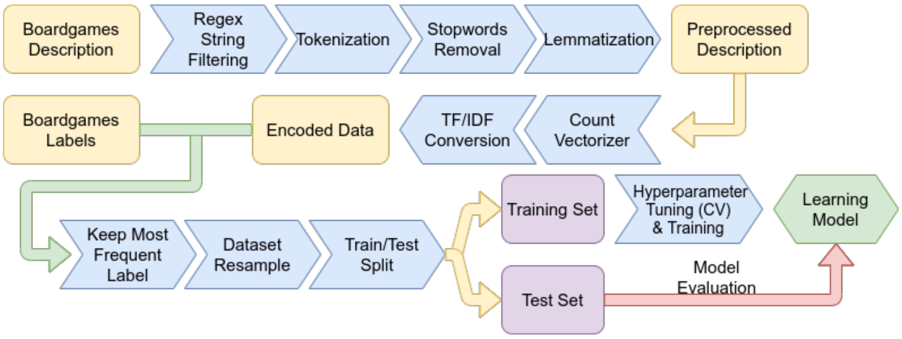

Fig. 1. Overall data processing and analysis pipeline

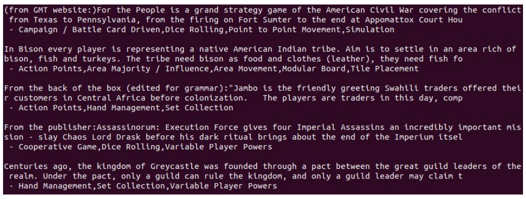

Fig. 2. Sample of 5 random records (raw data): game description (partial) and its associated game mechanics

manipulation, predictive model training, cross validation and test.

Dataset exploration. From Figure 2 we can observe excerpts of raw data, with descriptions and game mechanics. Through a first exploration, we noticed several aspects to be managed, including: (a) textual descriptions have to be pre-processed and cleaned in order to derive useful textual features (see next subsection); (b) as shown by the histograms (Figures 3 and 4), a strong unbalancing between the classes that would make the models' training less effective. The latter problem will be dealt with in the filtering and resampling phase, described later.

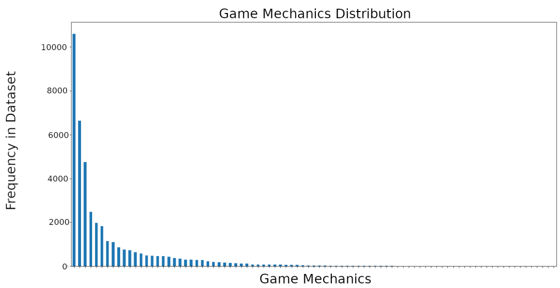

Fig. 3. Game mechanics distribution in the 50000 board games dataset

Text pre-processing and encoding. We implemented a simple pipeline of processing operations that performs that cleaning in a simple, compact and reusable way (see upper part of Figure 1). In particular, we perform the following steps: the text is cleaned (special characters removed and text converted to lower case) by means of a string filtering based on regular expressions matching, then the string is tokenized, the stopwords are removed and a lemmatizer is applied on the remaining

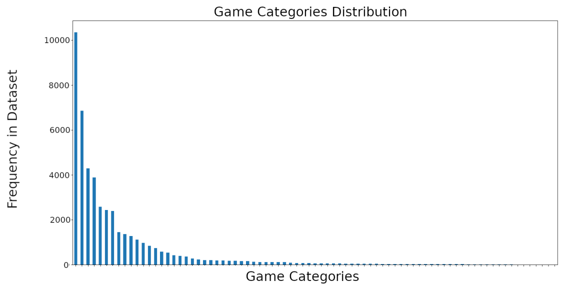

Fig. 4. Game categories distribution in the 50000 board games dataset

words by means of the WordNet thesaurus. Finally, we proceed with the encoding of the text descriptions in order to produce useful features for the models (see central part of Figure 1). The first stage of the encoding pipeline is the *count vectorizer* transformation which encodes the descriptions of the board games in a format that a learning algorithm can handle. This method creates a matrix W with dimensions $n \times m$ where n is the number of the different board games descriptions and m is the vocabulary composed of all different words found in them. Each element $w_{i,j}$ of the matrix W contains the number of occurrences of the word j in the $i-th$ description (*Term Frequency - TF - measure*). The second stage of the pipeline is meant to enhance the previous measure; the idea is that words appearing in most game descriptions will not give the model useful clues about game classification. Therefore, the final measure is obtained by multiplying TF by the *Inverted Document Frequency (IDF)*[6] of the term (TF-IDF weighting scheme), combining the frequency of a term in the description with the rarity (or specificity) of the term in the collection.

Data filtering and resampling. In the preliminary experiments we present in this paper, we focus on automatically determining the main mechanic (and main category) for a game. In order to solve the strong unbalance of the datasets, where there was a very small number of mechanics (and categories) representing the majority of the games, we proceeded with dataset filtering (keeping only the most fre-

[6]Defined as $IDF(t) = \log \frac{n}{df(t)}$ where t represents the term, n the total number of board games descriptions and $df(n)$ the number of board games descriptions where the term t appears.

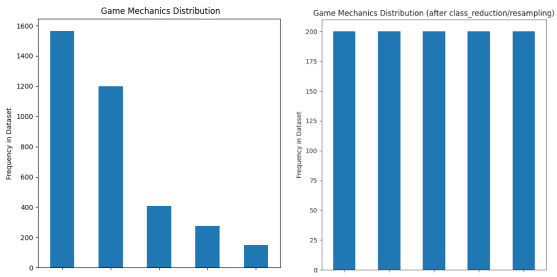

Fig. 5. Data resampling: unbalanced vs balanced mechanics

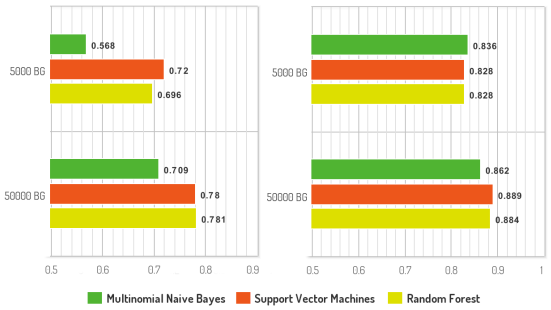

Fig. 6. Mechanics (left) and Categories (right) classification: accuracy of the models over the two datasets

quent mechanics/categories) and resampling (undersampling the most frequent mechanics/categories in order to balance the dataset and make the future learning of the models more effective). Figure 5 shows an example by considering the top 5 mechanics.

Models and hyperparameters tuning. The models we used for the experiments were chosen in such a way to test different strategies of learning:

- *Multinomial Naive Bayes (MNB)*: a classification technique based on the Bayes' theorem with an assumption of independence among predictors;
- *Support Vector Machines (SVM)*: one of the most robust prediction methods which is based on the Vapnik-Chervonenkis (VC) theory; it can be naturely extended to perform non-linear classification using the so called *"kernel trick"* which consists in mapping the inputs into a high-dimensional feature space;
- *Random Forest (RF)*: an ensemble method which relies on multiple decision trees whose outputs are used to form the final prediction using different strategies.

Every model has its own hyperparameters, including *C* (regularization parameter, range [0.5-4.0]) and *Kernel* (nonlinear separator hyperplane) for SVM, *Max_Est* (forest extension, range [100,1000]) for RF. Moreover, we defined some common hyperparameters relative to text processing:

- *ngram*: $\{(1,1), (1,2)\}$ a value of (1,1) means that the textual features are formed by single words, (1,2) means to exploit also couples of words;
- *max_feats*: $\{250, \ldots, 1000000\}$ only the *max_feats* most frequent features are kept;

	Method	Samp.	ngram	max_feats	use_IDF	C	Kernel	Max_Est
0	Multinom. NB	5K	(1..2)	2500	Yes	---	---	---
1	SVM	5K	(1..2)	2500	Yes	1.0	RBF	---
2	Random Forest	5K	(1..2)	4000	Yes	---	---	1000
3	Multinom. NB	50K	(1..2)	200000	Yes	---	---	---
4	SVM	50K	(1..2)	100000	No	2.5	RBF	---
5	Random Forest	50K	(1..2)	10000	No	---	---	1000

TABLE II
BEST HYPERPARAMETERS FOR GAME MECHANICS CLASSIFICATION

	Method	Samp.	ngram	max_feats	use_IDF	C	Kernel	Max_Est
0	Multinom. NB	5K	(1..2)	30000	Yes	---	---	---
1	SVM	5K	(1..2)	30000	Yes	2.0	Linear	---
2	Random Forest	5K	(1..2)	20000	No	---	---	1000
3	Multinom. NB	50K	(1..2)	400000	Yes	---	---	---
4	SVM	50K	(1..2)	60000	Yes	3.0	RBF	---
5	Random Forest	50K	(1..2)	60000	Yes	---	---	1000

TABLE III
BEST HYPERPARAMETERS FOR GAME CATEOGORY CLASSIFICATION

- *use_IDF*: $\{$True, False$\}$ if true, TF-IDF is exploited instead of TF.

The actual tuning was performed for each of the experiments through a *cross validation* stage using all possible combinations of hyperparameters and returning the best hyperparameter combination found. The cross validation strategy is the *K-Fold* one, where $K = 5$.

V. Obtained Results and Model Interpretation

In this section we detail the obtained accuracy results for mechanics and categories classification (tests performed for top 5 mechanics and categories).

Game mechanics classification. As shown by the histogram in Figure 6, the obtained accuracies are generally good, with all the three models reporting similar scores for both the small (top part of figure) and large (bottom part) dataset. SVM and RF are the methods which obtained the best scores (with a maximum of 78.1% for RF). We can also note the performance improvement obtained by using the large dataset. Moreover, looking at the chosen hyperparameters (shown in Table II) we can see that: (i) all the models benefited from features composed also of couple of words; (ii) for the dataset of 50000 board games, SVM and Random Forest have not benefited of the TF-IDF weighting scheme, disabling IDF; (iii) looking at the SVM kernel we have a clue about the non linearity of the datasets: the kernel used is the Radial Basis Function (or Gaussian Kernel) exploiting a nonlinear separator hyperplane.

Game categories classification. In this test, all classifiers obtained higher scores compared to the game mechanics classification task (Figure 6), in particular SVM scored an accuracy of 88.9% for the large dataset (versus 78% on the small one). Using the large dataset all three models obtained very high scores, especially if we consider that only the text description field was used as input.

The chosen hyperparameters (Table III) confirm many of the choices made for the previous task; notice that the models generally used a lower number of features with the exception

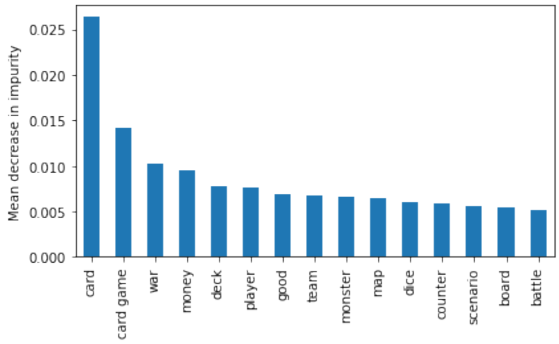

Fig. 7. Top 15 features ordered by importance (MDI)

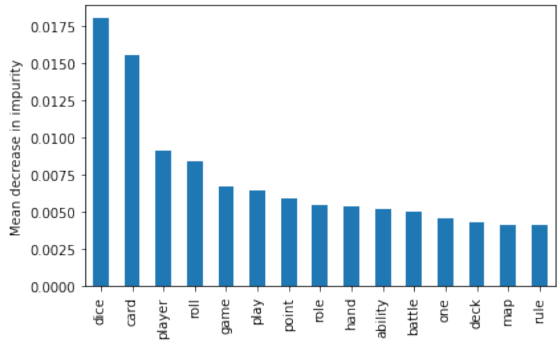

Fig. 8. Top 15 features ordered by importance (MDI)

of the MNB model with the larger dataset. As in the previous experiment, the IDF was often used but an important difference is the best kernel type selected for SVM model (small dataset) is linear, a clue that classifying categories is possibly a simpler problem than classifying mechanics. This hypothesis would be supported by the higher and not so different between 5K and 50K samples accuracies obtained, and by the lower increment of features selected from 5K to 50K with the respect to the game mechanics test.

Models interpretation. An important aspect of the proposed approach is to be able to gain insight on the machine learning processes and on the features that most influence it. This can be achieved through feature importance evaluation (or, for black-box models, through the SHAP[7] game-theoretic approach). In this preliminary work, we give an example of the features that most influenced the RF model for the classification tasks described in the previous section. As to game categories, by exploiting the transparency of the tree-based ensemble methods, we extracted the top 15 important features with their overall degree of influence (expressed in percent) for the top 5 categories (see Figure 7), including: "card", "card game" (*Card Game*), "money", "good", "counter" (*Economic*), "deck", "monster", "scenario", "board" (*Fantasy*), "player", "team", "battle" (*Fighting*), "war", "team", "counter" (*Wargame*). The same analysis has been carried out for the game mechanics model (Figure 8), where we got "card", "hand" and "deck" for mechanics such as *Card Drafting* and *Hand Management*, "point" for *Set Collection*, etc.

VI. Conclusions and Future Work

In this work, representing the first steps of this research, we collected a large dataset of board game information and explored the application of data analytics and machine learning techniques for the important and little explored objective of discovering "hidden" game features. The first tests show encouraging results. In the future, we plan to extend the models by considering additional features, for whose extraction we will also consider semantic techniques derived from the ones we employed in the past in different scenarios, from knowledge-based software engineering [13] to personalized medicine [8] and cultural heritage [5], [6]. Moreover, starting from additional inputs given by game experts on specific games, we plan to apply our data-driven approach to expand the scale of their intuitions by discovering currently unknown (but useful) aspects of games, e.g., the features that could foster Computational Thinking in education. The interpretability of the approach will be especially useful for such future analyses, as it could help in the generalization of interesting findings to be validated by the experts.

[7]SHAP, http://github.com/slundberg/shap

Acknowledgment

The authors would like to thank Andrea Ligabue and Marcello Missiroli for the stimulating discussions on game science in projects related to this research; Luca Giovannoni and Simone Dattolo, who contributed to the related work analysis and to the dataset extraction during their bachelor and master degree internships.

References

[1] Muhammad Aurangzeb Ahmad, Carly Eckert, Ankur Teredesai, and Greg McKelvey. Interpretable Machine Learning in Healthcare. *IEEE Intelligent Informatics Bulletin*, 19(1):1–7, 2018.

[2] Fabiola Bertolotti, Tommaso Fabbri, Federica Mandreoli, Riccardo Martoglia, and Annachiara Scapolan. Work datafication and digital work behavior analysis as a source of social good. In *IEEE Consumer Communications and Networking Conference (CCNC)*. IEEE, 2020.

[3] Michael Cook, Simon Colton, Jeremy Gow, and Gillian Smith. General analytical techniques for parameter-based procedural content generators. In *IEEE Conference on Games, CoG 2019, London, United Kingdom, August 20-23, 2019*, pages 1–8. IEEE, 2019.

[4] Davide Ferrari, Giovanni Guaraldi, Federica Mandreoli, Riccardo Martoglia, Jovana Milic, and Paolo Missier. Data-driven vs knowledge-driven inference of health outcomes in the ageing population: a case study. In *4th International workshop on Data Analytics solutions for Real-LIfe Applications, co-located with EDBT/ICDT 2020 Joint Conference (DARLI-AP @ EDBT 2020)*, 2020.

[5] M. Furini, F. Mandreoli, R. Martoglia, and M. Montangero. A predictive method to improve the effectiveness of twitter communication in a cultural heritage scenario. *ACM J. Comput. Cult. Herit.*, 2021.

[6] Marco Furini, Federica Mandreoli, Riccardo Martoglia, and Manuela Montangero. The use of hashtags in the promotion of art exhibitions. In *Digital Libraries and Archives - 13th Italian Research Conference on Digital Libraries (IRCDL), Revised Selected Papers*, pages 187–198, 2017.

[7] Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. In *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018*, pages 80–89, 2018.

[8] F. Grandi, F. Mandreoli, and R. Martoglia. Issues in personalized access to multi-version xml documents. In *Open and Novel Issues in XML Database Applications: Future Directions and Advanced Technologies, Eric Pardede (Ed.), IGI Global*, pages 199–230. 2009.

[9] Singh Kohli. Predicting Board Game Reviews using KMeans Clustering & Linear Regression. https://guneetkohli.github.io/machine-learning/board-game-reviews/#.YD1oo2hKjIU, 2016.

[10] Wolfgang Konen. General board game playing for education and research in generic AI game learning. In *IEEE Conference on Games, CoG 2019, London, United Kingdom, August 20-23, 2019*, pages 1–8. IEEE, 2019.

[11] Jakub Kowalski, Radoslaw Miernik, Maksymilian Mika, Wojciech Pawlik, Jakub Sutowicz, Marek Szykula, and Andrzej Tkaczyk. Efficient reasoning in regular boardgames. In *IEEE Conference on Games, CoG 2020, Osaka, Japan, August 24-27, 2020*, pages 455–462. IEEE, 2020.

[12] Joshua Kritz, Eduardo Mangeli, and Geraldo Xexéo. Building an Ontology of Boardgame Mechanics based on the BoardGameGeek Database and the MDA Framework. *SBGames*, 2017.

[13] R. Martoglia. Facilitate it-providing smes in software development: a semantic helper for filtering and searching knowledge. In *Proceedings of the 23rd International Conference on Software Engineering & Knowledge Engineering (SEKE'2011), Eden Roc Renaissance, Miami Beach, USA, July 7-9, 2011*, pages 130–136. Knowledge Systems Institute Graduate School, 2011.

[14] Francisco J. R. Ruiz, Christian A. Naesseth, Antonio Hernández, Manuel Soler, and Pedro Sánchez. Board Game Finder. https://www.boardgamefinder.net/, 2018.

[15] Robin van Gerwen. *Exploring the Relationship between Motivation, Mechanics and Genre for Tabletop Games*. PhD thesis, Tilburg University, 2019.

[16] Phil Woodward and Sam Woodward. Mining the BoardGameGeek. *Significance*, 16:24–29, 2019.