Discovering Critical Factors in the Content of Crowdfunding Projects

Kai-Fu Yang [1], **Yi-Ru Lin** [2] **and Long-Sheng Chen** [2,*]

[1] Department of Applied English, Chaoyang University of Technology, Taichung City 413310, Taiwan
[2] Department of Information Management, Chaoyang University of Technology, Taichung City 413310, Taiwan
[*] Correspondence: lschen@cyut.edu.tw

Abstract: Crowdfunding can simplify the financing process to raise large amounts of money to complete projects for startups. However, improving the success rate has become one of critical issues. To achieve this goal, fundraisers need to create a short video, attractive promotional content, and present themselves on social media to attract investors. Previous studies merely discussed project factors that affect crowdfunding success rates. However, from the available literature, relatively few studies have studied what elements should be involved in the project content for the success of crowdfunding projects. Consequently, this study aims to extract the crucial factors that can enhance the crowdfunding project success rate based on the project content description. To identify the crucial project content factors of movie projects, this study employed two real cases from famous platforms by using natural language processing (NLP) and feature selection algorithms including rough set theory (RST), decision trees (DT), and ReliefF, from 12 pre-defined candidate factors. Then, support vector machines (SVM) were used to evaluate the performance. Finally, "Role", "Cast", "Merchandise", "Sound effects", and "Sentiment" were identified as important content factors for movie projects. The findings also could provide fundraisers with suggestions on how to make their movie crowdfunding projects more successful.

Keywords: crowdfunding; natural language processing; text mining; feature selection; support vector machines

1. Introduction

When startups need capital in traditional financial markets, they often turn to financial institutions [1], but the cumbersome procedures and requirements have deterred startups. Crowdfunding has become an important source of investment capital for startups [2]. Global crowdfunding has grown by more than 33% in the last two years, with a large proportion of investors setting up fundraising projects for the first time [3]. Corresponding rewards come in the form of money contributed by investors [4,5], so startups have new opportunities to raise significant capital to complete projects in the program.

Relevant studies have found that the number of crowd fundraising cases in Taiwan has increased rapidly in recent years. However, statistics show that the number of projects and the amount sponsored have increased year by year, but the fundraising success rate has not increased. How to improve the fundraising success rate to complete the project has become one of the most important tasks for all fundraisers [6,7].

Fundraisers can simplify the funding process to achieve higher profitability through crowdfunding, and investors will take an active interest in investment projects [8] and then pay attention to the progress of the project, so that the project can receive more attention. To increase the success rate of fundraising, fundraisers need to create a short video that attracts investors, clear and attractive promotional content, and must maintain a presence through social media when writing a project [9].

In the traditional movie industry, there is public sponsorship and funding, etc., because large amounts of money are needed at each stage of production. After crowdfunding

became a new channel for fundraising, many producers launched projects to raise funds [8]. Previous research on movie crowdfunding has only focused on the project characteristics that affect the success rate of crowdfunding [7]. In the work of Lin and Boh [10], they indicated that the project description can influence the success of crowdfunding projects. However, few studies have taken an in-depth look at the project content and informed fundraisers about what elements should be involved to increase the success rate. Chen et al. [7] defined factors for project characteristics, descriptions, and evaluations of movie projects but focused less on the factors of project content descriptions. Chen et al. [11] discussed the factors that need to be included in the project content for a successful project, based on the topic of game projects.

In addition, previous research mainly relied on quantitative research based on questionnaires. Other than the labor and time required for data collection, it is also prone to sampling error. Schuckert et al. [12] stated that traditional questionnaire surveys are prone to experimental effects and the information obtained through online text content is more objective, massive, and less prone to sampling bias than the use of questionnaires. In addition, related research attempted to use natural language processing (NLP) and text mining from social media text comments or online text data to understand customer voices [6,13].

Therefore, based on Chen et al. [7], this study observed possible elements for the content description of movie projects and further searched for factors that influence the success of movie fundraising projects. This study aimed to discover the key factors of movie crowdfunding project content, and then used text mining and feature selection approaches, including rough set theory (RST), decision trees (DT), and ReliefF methods to select important candidate factors. The subsets of candidate features selected by the feature selection method are then evaluated by support vector machines (SVM). This study identified the key factors that affect the success of crowdfunding for movies and can provide useful recommendations for future crowdfunding to improve the success rate of crowdfunding.

2. Related Works

2.1. Potential Factors for Successful Movies

According to the existing literature, there are a few factors that affect the definition of the problem of crowdfunding for movies. Therefore, this study set out to search the movie-related literature for factors that influence the success of movies as potential candidates for influencing crowdfunding projects for movies.

There are many recent studies on the success factors of movies, such as Verma and Verma [14], who found that almost no movie can win the appreciation of all audiences. Mina and Baber [8] showed that actors, scripts, distributors, financing, and merchandising are the key elements for a successful movie. Wang et al. [15] pointed out that the success of a movie depends on factors, such as actors, directors, scripts, shooting skill, social media advertisement, and box office revenue. Shooting requires three main participants, namely the actor, the director, and the production company. Their performances are considered an important factor in the ultimate box office revenue. It mentioned that most movie experts believe that a movie's story, script, or screenwriter can predict how the final movie will perform at the box office. As we all know, the more people enter the theater, the higher the box office revenue. For this reason, movie promotion usually involves advertising campaigns by the distribution company to make the audience aware of the movie and arouse their curiosity. Kang et al. [16] believed that advertising, word of mouth, star power, online media rating, online media popularity, and industry recognition are important factors. Wei and Yang [17] indicated that budget and producer are very important for a successful movie. Zhang and Zhang [18] specifically pointed out that the story, plot, actors, ending, acting, director, era, rhythm, image, setting, character, male lead, screenwriter, soundtrack, details, original work, theme, special effects, style, lines, logic, background, photography, and opening are all relevant and crucial. The research results of Moon

et al. [19] showed that script, box office, revenue, budget, sequel, genre, era, distribution, star power, and language are all important to the success of a movie. Table 1 summarizes success factors of a movie in related works.

Table 1. Success factors of a movie.

Related Literature	Success Factors
Mina and Baber [8]	Actor, script, distributor, funding, merchandise
Wang et al. [15]	Actor, director, script, shooting skills, social media advertisement, box office revenue
Kang et al. [16]	Advertisement, word of mouth, star power, online media evaluation, online media popularity, industry recognition
Wei and Yang [17]	Budget, producer
Zhang and Zhang [18]	Story, plot, actor, ending, acting, director, era, rhythm, picture, shot, character, male lead, screenwriter, soundtrack, details, female lead, original work, subject matter, special effects, style, lines, logic, background, photography, beginning
Moon et al. [19]	Script, box office performance, revenue, budget, sequel, genre, year, publisher, star power, language

2.2. Text Mining

Text mining aims to recognize the important information in documents, and to discover useful information from the document. In other words, text mining is a method for processing, organizing, and analyzing a huge number of documents. Its primary objective is to turn text into data for natural language processing [20].

Nowadays, the size of information from the Internet grows dramatically. These huge amounts of unstructured or semi-structured text data need to be processed using text mining techniques to find out the hidden structures and rules [21]. Turban et al. [22] divided the text mining process into three steps, namely building a corpus, creating terms (through a term-document matrix, TDM), and extracting knowledge or patterns.

Some studies adopted text mining to find potential features of Chinese characters for crowdfunding project presentations. In addition, Wang et al. [23] analyzed the influence of project descriptions and project founders' emotions on crowdfunding success rate. Du et al. [24] investigated the quality and source credibility of crowdfunding project descriptions and analyzed the impact on crowdfunding project success.

Text mining has also been successfully utilized in many areas. For examples, Loureiro [25] conducted a full-text analysis of VR and AR journals and conference proceedings by using text mining methods. In the work of Zhong et al. [26], they integrated deep learning into text mining to analyze hazard construction data from unstructured or semi-structured documents.

In summary, text mining is frequently utilized and has a high level of success in analyzing unstructured texts. As a result, the text mining method will be employed to deal with the text content of the project description. Moreover, we will build lexicons for each factor to construct our experimental data.

2.3. Feature Selection

Feature selection aims to reduce irrelevant data in the original feature set and obtain discriminative and effective feature subsets to achieve the effect of dimensionality reduction and determine the best feature subset to achieve the best process [27].

2.3.1. Rough Set Theory (RST)

Rough set theory (RST) was developed by Pawlak [28] as a solution for confusing or vague notions. It is mostly used for learning and summarizing incomplete facts and ambiguous information. It is also suitable for determining the potential connection between

knowledge and discovery and for identifying significant critical components [29,30]. Finding a set of attribute subsets that are less than the original attributes is the basic goal of RST. This attribute subset shares the same categorization capabilities as the original attribute set, a process known as attribute reduction. If eliminating duplicate features does not decrease the classification accuracy, we can locate a more condensed collection of attributes [31].

Rough set theory has been used extensively in machine learning, knowledge discovery, data mining, spam filtering, gene expression analysis, classification tree induction, and feature selection [32–34]. The chosen subset can be used for classification, regression, clustering, outlier detection, and other learning algorithms, such as decision trees, naive Bayes, the support vector machine, and k-nearest neighbor, to evaluate its effectiveness [35].

2.3.2. Decision Trees (DT)

Decision trees is widely used in classification, prediction, and other areas. A DT is usually utilized as a prediction model. However, when a DT has been considered as a feature selection tool, all attributes that appear in the built trees will be considered important [6]. In this study, we employed DT to be a feature selection tool.

There are many successful cases of applying decision trees. For example, Chang et al. [6] used decision trees (C5.0), least absolute shrinkage and selection operator (LASSO), and support vector machine recursive feature elimination (SVM-RFE) to determine the critical features influencing customers' non-revisit intentions. In the work of Rao et al. [36], they presented a new feature selection algorithm based on bee colonies and gradient boosting decision trees. Their results confirmed that the proposed algorithm can reduce the dimensionality and achieve an outstanding performance. In summary, this study also uses decision trees to be one of the feature selection methods.

2.3.3. ReliefF

In the relevant literature, ReliefF has shown its powerful functions, and there have been many successful applications in real world. For examples, Shi et al. [37] developed a highly efficient fault diagnosis model, which used the ReliefF algorithm for feature ranking and successfully applies neural networks for variable refrigerant flow system fault diagnosis. Jin et al. [38] proposed a new method based on ReliefF-SVM to study the potential relationship between Parkinson's disease and scans without evidence of dopaminergic deficit. In addition, Aslan et al. [39] used ReliefF to improve the performance of a trained deep convolutional neural network (CNN) and used it to help radiologists diagnose COVID-19 early and automatically on X-ray images. Wen and Ziang [40] utilized ReliefF for feature selection and used a deep neural network (DNN) to build a wind farm fault judgment module. This method can accurately diagnose wind turbine faults.

Furthermore, in the work of Kilicarslan et al. [41], they employed ReliefF for dimensionality reduction, and support vector machines (SVM) and convolutional neural networks (CNN) for classification. From their experiments on three microarray datasets of ovary, leukemia, and the central nervous system, it can be found that the dimensionality reduction method improves the classification accuracy of the SVM and CNN methods. Souza et al. [42] utilized machine learning to predict cadmium concentrations in plants using kale (Brassica oleracea) and basil (Ocimum basilicum). In their model, ReliefF has been employed as feature selection method. Zhang et al. [43] proposed a new random multi-subspace-based ReliefF for feature selection and verified it with 28 UCI datasets of different sizes. Their results proved that effectiveness of their ReliefF based method in solving feature selection problems. Consequently, this study also employs ReliefF to be one of the feature selection methods.

3. Methodology

This section will describe the methodology employed in this study. The implementation process can be divided into 8 steps, which are described in detail as follows.

Step 1: Collect data

The fundraising projects used in this study comes from the work of Chen et al. [8]. The crawler program is written in Python language to extract project content parts from two famous crowdfunding platforms, namely "Indiegogo (https://www.indiegogo.com/ accessed on 1 July 2022)" and "Kickstarter (https://www.kickstarter.com/ accessed on 1 July 2022)". We select the top three types of movie projects, which could be divided into "comedy", "narrative movie", and "drama", with the highest project success rate ranking.

Step 2: Define Candidate Content Factors

From the available literature, very few studies mention the key content factors in the success of movie crowdfunding projects. Therefore, this study uses the keywords "crowdfunding" and "movie success" to search related studies to find candidate factors for movie project content. Finally, we define 12 factors that affect the success of movie project content, as shown in Table 2.

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

Table 2. Candidate content factors affecting the success of movie projects.

No.	Factor	Definition	Supports
X1	Storyline	The project content mentions movie storyline, such as plot, background, script, etc.	[11,17–19,38,39]
X2	Role	The project content mentions movie characters, such as characters, roles, characteristics, occupations, etc.	[11,17,18]
X3	Cast	The project content mentions the film cast, such as actors, directors, production companies, etc.	[7,16,38]
X4	Merchandise	The project content mentions movie peripheral products, such as merchandise, commemorative merchandise, clothing, movie soundtracks, etc.	[8,11,39]
X5	Advertisements	The project content mentions traditional advertising for movies, such as TV ads, magazine ads, station commercials, etc.	[11,38,39]
X6	Social media	The project content mentions the social media marketing of the film, such as Facebook, Twitter, YouTube, Instagram, and other communities.	[7,11,40]
X7	Funding	The project content mentions the funding of the film, such as sponsors, total budget, cost, etc.	[7,17,19]
X8	Screen features	The project content mentions the features of the movie's screening, such as 3D, scenes, animations, etc.	[11,17,18,39]
X9	Sound effects	The project content mentions sound effects, such as classical, musical instruments, stereo effects, etc.	[8,11,18]
X10	Positive sentiment	The project content contains positive sentiment.	
X11	Negative sentiment	The project content contains negative sentiment.	[7]
X12	Sentiment (positive–negative)	The overall sentiment of the project content text (positive sentiment-negative sentiment).	

Step 3: Build the Lexicons

To define the 12 content candidate factors, a lexicon will be built for each factor in this step. This study uses the synonym website "Thesaurus.com" (https://www. thesaurus.com/ accessed on 1 February 2022) to define its similar thesaurus for X1 to X9. In addition, the lexicons of three sentiment factors (X10~X12) will used SentiWordNet (http://sentiwordnet.isti.cnr.it/ accessed on 1 February 2022).

Step 4: Construct Experimental Data

In the past, NLP used terms as attributes to establish a term–document matrix (TDM). In this study, lexicons are used as attributes. The content description of each project is compared with the established lexicons, and the frequency of occurrence of related words is counted. Then, we can build experiment data, namely "Lexicon-Document Matrix (LDM)". The detailed steps have been provided as follows:

(1)   Use the built lexicons in step 3;
(2)   Text preprocessing, including deletion of symbols, stop words, etc.;
(3)   Stemming;
(4)   English word segmentation (this study adopts the unigram method);
(5)   Word frequency statistics;
(6)   Establish experimental data.

Step 5: Implement Feature Selection

In order to reduce the error caused by the experiment, the experimental data is normalized into the interval [0, 1]. After normalization, a five-fold cross-validation method is used to divide the data into five equal parts, which are the test and training data sets, respectively, and to cross-validate the test data set and training data set in sequence, to make the results more accurate and reliable, and screen important feature subsets according to the following feature selection methods. In this study, we employed three feature selection methods.

3.1. Decision Trees

In this study, a DT has been utilized to perform the feature selection. First, we build a tree by using the C5.0 algorithm. Then, the node left in the constructed tree will be considered as important. The implemental steps are as follows:

Step 1 Define the input and output factors;
Step 2 Construct DTs for each fold data set;

 Step 2.1 Create an initial rule tree;
 Step 2.2 Prune this tree;
 Step 2.3 Process the pruned tree;

Step 3 Determine the important factors from built trees.

3.2. Rough Set Theory

The RST approach handles information represented by the information system containing samples and features. In feature selection, RST implements the attribute reduction process of finding an optimal attribute subset with the same or better classification performance. The reduced feature subset is called "reduct", which is the essential part of an information system. By means of the dependent properties of the attributes, we can remove the redundant features to find the optimal feature subset [31].

3.3. ReliefF

The third used feature selection method is ReliefF [37], which is an extension of Relief. ReliefF is more powerful than the original one and can handle multiple classes of problems. ReliefF uses k-nearest hits and nearest misses, and then updates to the predicted vector quality of attribute A after averaging. Actually, this method is a feature weighting algorithm, which assigns different weights according to the correlation between each feature and the class. First, randomly select a sample R from all samples, and then take out the nearest neighbor samples I_1 and I_2 from the sample group of the same classification as sample R. One is nearest hit H in the same class as R, and the other is nearest miss M in a different class. Update W [A] for all attributes and repeat m times. The algorithm is as follows [37]:

```
            Input: attribute vector values and class values for each training instance
            Output: Predicted vector quality for attribute A
      Set all weights W [A]:=0.0;
      For i: =1 to m do begin
            Randomly select an instance R;
            Find nearest hit H and nearest miss M;
            For A: =1 to #all_attributes do
                  W [A]:=W [A] − diff (A, R, H)/m + diff (A, R, M)/m;
      End;
```

In addition, diff (A, R, H) is defined for discrete attributes as shown in Equation (1). *A* is an attribute, and I_1 and I_2 are samples. Equation (1) is as follows:

$$diff(A, I_1, I_2) = \left(\begin{array}{l} 0; value(A, I_1) = value(A, I_2) \\ 1; otherwise \end{array} \right. \tag{1}$$

Step 6: Evaluate the selected subset of factors by SVM

Based on the feature subset established in step 5, this study uses support vector machines for performance evaluation. The SVM classifier will be built separately from the reduced feature subset and the original feature set. If the reduced feature subset has a smaller number of features, it still can have a similar or even better accuracy. This means that these fewer factors are more informative and the selected factors could be considered as important.

Step 7: Identify key factors

After constructing SVM classifiers by feature subsets which were extracted by feature selection methods, we use overall accuracy (OA), F1-measure, and training time to confirm important factors. Based on the results of the analysis, recommendations will be given to movie project owners through crowdfunding. This study uses a confusion matrix to calculate the metrics, as shown in Table 3, consisting of TP, FP, TN, and FN. These abbreviations are defined as follows:

TP—predicted successful, actually successful;
FP—predicted Successful, actually failed;
TN—predicted failed, actually failed;
FN—predicted failed, actually successful.

Table 3. Confusion matrix.

Predicted Actual	Predicted Positive (Successful)	Predicted Negative (Failed)
Actual Positive (Successful)	TP (True Positive)	FN (False Negative)
Actual Negative (Failed)	FP (False Positive)	TN (True Negative)

According to the above confusion matrix, OA and F1 are calculated as follows in Equations (2) and (5), respectively:

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{Precision + Recall} \qquad (5)$$

Step 8: Draw Discussion and Conclusion

Based on the results of the experiment, we identified important content factors that influence the success of a movie project. Based on this result, we can provide specific advice to movie project fundraisers on how to write a successful project description.

4. Implementation

4.1. Employed Data

The data for this study is from Chen et al. [7], which collects data from the crowdfunding sites "Indiegogo" and "Kickstarter" on the introductory part of the movie project. The original data study collected the top three types of movie projects, which are "comedy", "narrative movie", and "drama", ranked by the highest success rate of movie projects. Projects that "raised 100% or more of the target amount" were defined as successful projects, while projects that "raised between 0 and 100% of the target amount" were considered failures. Moreover, in this study, after deducting the data with random numbers and no text displayed in the content description of the collected projects, the employed data for further study is shown in Table 4. Figure 1 provides an example of a movie project on the Kickstarter crowdfunding platform. We only utilized the text description part of the project content in this figure.
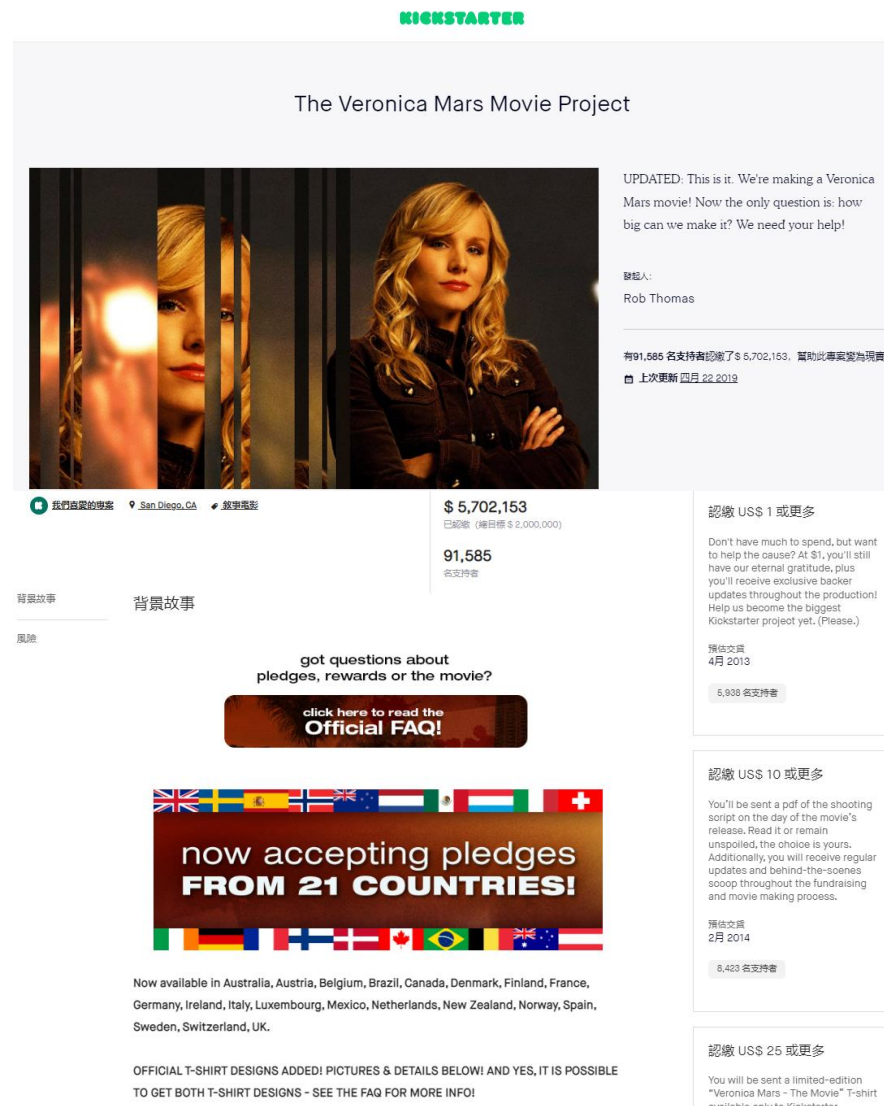
Figure 1. An example of movie projects in Kickstarter crowdfunding platform.

Table 4. Employed data of this study.

Fundraising Platform / Data Size	Indiegogo	Kickstarter
Success projects	297	1014
Failed projects	646	429
Total	943	1443

4.2. Defining Candidate Factors and Establishing Lexicons

Since there is no study to discover the success content factors of movie projects from the available literature, this study will use "key success factors of crowdfunding projects" and "movie success factors" [7,8,16–19,38–40] as our candidate factors.

After surveying the available literature, we defined 12 candidate factors. Next, we must create lexicons for every single factor according to the definition of the factors. Table 5 lists the 12 candidate factors and examples of the related lexicons. In this table, with respect to factors X1 to X9, we used the synonym website "Thesaurus.com" to build a relevant representative dictionary for each factor. For the semantic parts of X10 to X12, we use the sentiment lexicon website, SentiWordNet, as our lexicons to determine sentiments.

Table 5. The candidate factors of movie project content and their lexicons.

No.	Factors	Examples of Constructed Lexicons
X1	Storyline	Device, scenery, profile, outtake, unflinching, blessing, goods, status, attainment, notice, law, cover, lineage, archive, bulk, persistent, boards, procedure, preprint, invention, adventures, fake, remuneration, dividend, humor, blurb, gamesmanship, inscription, factor, envelope, card, schoolwork, incident, platitude, heritage, misadventure, bunk, shovel, epilogue, plot, curtains, narrative, recountal, performance, thread . . . . . .
X2	Role	Surrogate, crux, wrinkle, partition, member, system, vestment, labor, transaction, rubout, awarding, flake, pretense, constitution, holdall, apportionment, usage, stripe, bestowal, lineation, curve, pomp, masterpiece, gusto, administration, warmth, stratagem, slice, enterprise, object, tincture, demarcation, generosity, band, fiber, enlistment, achievement, product, dress, boldness, patronage . . . . . .
X3	Cast	Lucy Pinder, Peter Weingard, Lawrence Olivier, Joshua Jackson, Dakota Goyo, Max Mingra, Jordan Prentice, William Shatner, Patrick Adams, Megan Orly, Mia Kirchner, James Fox, Victor Jabo, David Warner, Ophelia Ravibond, Paul Gross, Robert Kasinski, Jim Stegers, Eugene Levy, Tracey Spiridacos, Sean Biggerstaff, Alyssa Nicole Pallett, Kristen Bell, Jack Houston, Ned Sparks . . . . . .
X4	Merchandise	Hike, run-of-the-mill, vendibles, blemished, rule, profile, honor, amiss, exaltation, improvement, output, character, wares, encouragement, line, streak, vendible, products, furrow, overused, worldly, outgrowth, impaired, set, unhealthy, by-product, objective, concrete, upshot, tracing, borderline, truck, flawed, nonspiritual, actual, demarcation, compound, result, digit . . . . . .
X5	Advertising	Bulletin, biweekly, boost, convolution, coil, advancement, brochure, annunciation, adjustment, conviction, aperçu, annular, brief, colloquial, broadside, communal, antagonism, carnival, chest, conjunction, beneficiary, break, communion, bung, belles-lettres, avenue, complect, confederation, architecture, bimonthly, conflicting, classification, cool, advisory, bulldog, coherence, cork, conversation, cord, bill, annual, converse, confirmation . . . . . .

Table 5. *Cont.*

No.	Factors	Examples of Constructed Lexicons
X6	Social media	Google Mail, camper, Viber, mobile home, QQ, house trailer, caravan, social platform, Tumblr, prevue, promo, teaser, Reddit, LinkedIn, camp trailer, doublewide, YouTube, trailer, social media service, Discord, WeChat, motor home, social networking website, Line, mail, social media website, Twitter, recreational vehicle, Tumbler, IG, Weibo, Pinterest, social media platform, Qzone, FB, RV, trail car, social media, Instagram, Telegram, Snapchat, Facebook, Quora, website, Vk, WhatsApp . . . . . .
X7	Funding	Mother, bestowal, confidence, dull, marketing, groundwork, banal, fundamental, lagoon, heart, antecedent, parent, cardinal, granary, chest, property, leading, reliance, customary, informant, nature, lot, cause, bank, archive, prime, line, garner, capital, pool, reason, income, normal, guts, natatorium, depository, mine, file, outstanding, gratuity, expert, provenance, onset, infrastructure . . . . . .
X8	Screen features	Class, oomph, sprightliness, penumbra, armament, fury, depiction, mien, customary, collateral, aura, miscellaneous, modicum, penchant, hardiness, fashion, gyration, hilarity, resilience, lump, outdoors, exhibition, absorber, mantle, maturity, dissimulation, ostentation, kidding, counterfeiting, drawing, curvilinear, litheness, consuetude, atom, prearranged, ritual, adroitness, stimulus, dumps, auditorium, spin . . . . . .
X9	Sound effects	Din, friendship, fusion, core, cacophony, blending, endeavor, carol, singing, angle, coalescence, bang, crooning, mellow, lot, constriction, direction, bit, communique, rap, reach, punch, courage, plasticity, scale, racket, approved, aim, row, bookish, round, cobblestone, societal, lodge, breeze, repercussion, state, motion, litany, pull, narrative, fraction, latest, emphasis, amalgam, family, motif, enhanced, magnitude . . . . . .
X10	Positive sentiment	Support, outstrip, geeky, adulation, agreeableness, soundly, diligently, congratulatory, nicest, gumption, immaculate, engaging, prefer, satisfy, luminous, unequivocally, restored, holy, protect, tops, ideally, insightfully, poeticize, wonderfully, adequate, rejoice, feat, courageously, cohesive, protection, acclamation, morality, astonished, preferring, long-lasting, excellent, marvelousness, securely, peaceable, contribution, homage, colorful . . . . . .
X11	Negative sentiment	Brutally, chintzy, disagreeably, despised, blab, dings, delay, conspiratorial, frantic, flickering, divisiveness, contempt, brutalizing, disgustingly, discordant, discriminate, fault, anxiously, forged, evils, drippy, dread, gall, fetid, bristle, anguish, craps, discontented, counter-productive, denigrate, disingenuously, hardliner, compulsion, bust, forceful, annoying, depression, abominably . . . . . .
X12	Sentiment (positive– negative)	Unwatchable, proper, integrated, impiety, problems, misgivings, trusty, shortsightedness, record-setting, inflated, divisive, mischief, proven, slumping, disintegration, obscure, cruelties, sensitive, problematic, genial, concerned, concede, trophy, resilient, tenderness, unspeakable, sensations, perturb, rubbish, spotty, dissatisfy, proves, cute, grumble, coherent, jubilantly, affirmative, intriguingly, unbearable, dissuasive, triumphal . . . . . .

4.3. Feature Selection

This study employs natural language processing (NLP) to deal with text data of movie project content. A five-fold cross-validation experiment has been performed. Regarding rough set theory, we use the package software Rosetta. In Rosetta, we first use entropy to discretize attributes' values to obtain the smallest feature subset (reduct). Then, we summarize the important feature subsets from the results of the five-fold experiments. Concerning the decision tree, the software See5 has been selected to execute the C5.0 algorithm. ReliefF is implemented using the Weka 3.8 software and uses preset cross-validation to evaluate the value of a given attribute by repeatedly sampling the data by considering recent data of the same and different classes. In terms of the decision tree, in C5.0, the pruning CF affects the way of estimating the error rate, thereby affecting the

severity of pruning, in order to avoid overfitting of the model. In this study, the pruning CF was set to 25%.

4.3.1. Results of Rough Set Theory

Table 6 shows the results of performing RST on the Indiegogo dataset. In the experiment for fold 2, three reducts of the same length are generated. Therefore, considering the experimental results of the other four folds, we use the frequency of occurrence to select important features. In this table, the factor with the highest frequency of seven times or more establishes the feature subset "RS-I1 {X3, X4, X6, X7, X10, X11}", and the frequency of five times or more is used as the feature subset "RS-I2 {X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12}". Table 7 shows the results of the Kickstarter data set. According to the frequency of occurrence, the feature subset "RS-K1{X3, X4, X5, X7, X8, X9, X10, X11, X12}" has been constructed.

Table 6. Factors selected by RST in the Indiegogo dataset.

Factor \ Fold	1		2		3	4	5	Frequency
X3	X	X	X	X	X	X	X	7
X4	X	X	X	X	X	X	X	7
X6	X	X	X	X	X	X	X	7
X7	X	X	X	X	X	X	X	7
X10	X	X	X	X	X	X	X	7
X11	X	X	X	X	X	X	X	7
X5		X	X	X	X	X	X	6
X12	X	X	X	X	X	X		6
X2	X		X		X	X	X	5
X8	X	X	X	X		X		5
X9	X	X			X	X	X	5
X1	X			X	X		X	4

Note, "X" represents the factor that was selected as an important factor in individual fold experiments.

Table 7. Factors selected by RST in the Kickstarter dataset.

Factor \ Fold	1	2	3	4	5	Frequency
X1	X	X	X	X	X	5
X2	X	X	X	X	X	5
X3	X	X	X	X	X	5
X4	X	X	X	X	X	5
X5	X	X	X	X	X	5
X7	X	X	X	X	X	5
X8	X	X	X	X	X	5
X9	X	X	X	X	X	5
X10	X	X	X	X	X	5
X11	X	X	X	X	X	5
X12	X	X	X	X	X	5
X6	X		X	X	X	4

Note, "X" represents the factor that was selected as an important factor in individual fold experiments.

4.3.2. Results of DT

Table 8 shows the results of the decision tree in the Indiegogo dataset. We established factors with a frequency of more than three times as a feature subset "DT-I1 {X12}". Table 9 lists the results in the Kickstarter dataset. We established the feature subset "DT-K1 {X2, X3, X4, X9, X12}".

Table 8. Factors selected by DT in the Indiegogo dataset.

Factor \ Fold	1	2	3	4	5	Frequency
X12	X	X		X		3
X5	X			X		2
X7		X		X		2
X9		X		X		2
X1	X					1
X3	X					1
X4		X				1
X8	X					1
X11	X					1
X2						0
X6						0
X10						0
Accuracy	62.4%	67.2%	68.3%	68.3%	69.5%	

Note, "V" represents the factor that was selected as an important factor in individual fold experiments.

Table 9. Factors selected by DT in the Kickstarter dataset.

Factor \ Fold	1	2	3	4	5	Frequency
X2	X		X	X		3
X3	X		X	X		3
X4	X		X	X		3
X9	X		X	X		3
X12	X		X	X		3
X1	X			X		2
X5	X			X		2
X6	X		X			2
X7	X			X		2
X11	X		X			2
X8	X					1
X10	X					1
Accuracy	68.5%	70.2%	68.9%	69.6%	70.4%	

Note, "X" represents the factor that was selected as an important factor in individual fold experiments.

4.3.3. Results of ReliefF

We use the parameter "ReliefF Attribute Eval" in Weka and use the cross-validation to calculate the average weight of each factor and rank it. Table 10 establishes a feature subset "RF-I1{X6, X7, X8, X10, X12}" for factors with a frequency of three or more in the Indiegogo dataset. Table 11 establishes a feature subset "RF-K1{X4, X6, X7, X8, X10, X11, X12}" for

factors with a frequency of three or more in the Kickstarter dataset. Table 12 summarizes the results of the three feature selection methods for the two different datasets.

Table 10. Factors selected by ReliefF in the Indiegogo dataset.

Fold / Factor	1	2	3	4	5	Frequency
			Rank			
X6	1.8	3.1	5.9		3.1	4
X7	7.1	6.2	2.5		5.1	4
X10	4.6	4.3	6.3		4.2	4
X12		5.5	3.8	1.1	3.8	4
X8	1.3	1.1	3.2			3
X4			5.2	2.1		2
X5				6.5	2.5	2
X9	5.8			4.3		2
X11	5.3				5.0	2
X1		3.1				1
X2				6.3		1
X3				3.8		1

Table 11. Factors selected by ReliefF in the Kickstarter dataset.

Fold / Factor	1	2	3	4	5	Frequency
			Rank			
X7	2.4	4.6		3.5	3.0	4
X8	4.7	3.5		2.3	5.0	4
X12	3.6		3.1	2.5	5.9	4
X4	6.8	4.7			3.0	3
X6			3.3	6.0	2.9	3
X10	2.9	5.9		3.2		3
X11	2.4	2.0	3.2			3
X1			3.2	6.5		2
X2			4.7			1
X3		6.0				1
X5					4.6	1
X9			5.9			1

4.4. Performance Evaluation by SVM

To evaluate the performance of a subset of extracted important attributes shown in Table 12, we will evaluate the with the SVM classifier and compare them with the original set of attributes. If the performance of the SVM classifier established with a smaller number of features can achieve similar performance to that of the SVM classifier established with all attributes, it means that these few attributes represent the same amount of information as all attributes. These few attributes can be identified as an important set of attributes.

Table 12. Summary of three feature selection methods in the Indiegogo and Kickstarter datasets.

Dataset	Feature Selection	Feature Set	Extracted Factors
Indiegogo	RST	RS-I1	X3, X4, X6, X7, X10, X11
		RS-I2	X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12
	DT	DT-I1	X12
		DT-OS-I1	X1, X2, X3, X4, X5, X7, X8, X9, X10, X11, X12
	ReliefF	RF-I1	X6, X7, X8, X10, X12
Kickstarter	RST	RS-K1	X1, X2, X3, X4, X5, X7, X8, X9, X10, X11, X12
	DT	DT-K1	X2, X3, X4, X9, X12
	ReliefF	RF-K1	X4, X6, X7, X8, X10, X11, X12

In this study, LIBSVM (Fan et al., 2005) is used to build the SVM prediction models, and we employ the C-SVC mode and radial basis function (RBF) kernel function. In addition, a parameter selection tool, grid.py, has been utilized to find the optimal parameter settings in LIBSVM.

Table 13 summarizes the feature subset evaluations for the Indiegogo dataset. From this table, the classification performance obtained by the feature subsets RS-I1, RS-I2, DT-I1, and RF-I1 selected by the three feature selection methods is compared with the original feature subsets. Considering F1 and training time, the performance of the RS-I2 feature subset is not far from the classifier performance of the original feature subset. As such, RS-I2 was selected as the best feature subset of the Indiegogo dataset.

Table 13. Summary of feature subset evaluation of the Indiegogo dataset.

Feature Set / Index	Original (12)	RS-I1 (6)	RS-I2 (11)	DT-I1 (11)	RF-I1 (5)
OA (%)	68.46 (0.45)	68.51 (0.57)	68.40 (0.67)	68.51 (0.57)	68.51 (0.57)
F1 (%)	1.30 (1.78)	0.00 (0.00)	0.63 (1.42)	0.00 (0.00)	0.00 (0.00)
Time (sec.)	0.04 (0.00)	0.03 (0.01)	0.04 (0.01)	0.03 (0.01)	0.05 (0.02)

Table 14 summarizes the feature subset evaluations for the Kickstarter dataset. In this table, from the perspective of OA, DT-K1 (69.99%) is slightly better than RS-K1 (68.89%) and RF-K1 (69.37%), and is similar to the 70.03% value of the original feature subset. If considering F1 and training time, it can be seen that DT-K1 (82.29%, 0.06 s) significantly outperforms RS-K1 (80.99%, 0.10 s), RF-K1 (81.52%, 2.86 s), and the original feature set (81.66%, 0.33 s). Therefore, this study selects the feature subset DT-K1 as an important feature subset in the Kickstarter dataset.

Table 14. Summary of feature subset evaluation of the Kickstarter dataset.

Feature Set / Index	Original (12)	RS-K1 (11)	DT-K1 (5)	RF-K1 (7)
OA (%)	70.03 (1.29)	68.89 (2.57)	69.99 (0.48)	69.37 (1.52)
F1 (%)	81.66 (0.92)	80.99 (3.01)	82.29 (0.43)	81.52 (1.85)
Time (s)	0.33 (0.29)	0.10 (0.04)	0.06 (0.01)	2.86 (3.77)

5. Discussions and Suggestions

From Table 13, it is easy to see that there are class imbalance problems, namely that the SVM classifiers built from the imbalanced dataset have a very low F1 in the Indiegogo dataset. Therefore, we implement SMOTE (synthetic minority over-sampling technique) for the Indiegogo data. Table 15 summarized the results.

Table 15. Summary of SMOTE results evaluation of the Indiegogo dataset.

Index / Feature Set	Original (12)	RS-I1 (6)	RS-I2 (11)	DT-I1 (11)	RF-I1 (5)
OA (%)	68.46 (0.45)	52.45 (7.85)	70.53 (9.82)	67.77 (7.17)	45.03 (8.47)
F1 (%)	1.30 (1.78)	53.03 (7.98)	68.46 (5.82)	39.25 (8.12)	45.91 (9.13)
Time (sec.)	0.04 (0.00)	0.08 (0.05)	0.08 (0.02)	0.07 (0.01)	0.07 (0.03)

Compared to Table 13, we can see that the F1 scores in Table 15 have a significant improvement. The overall accuracy also reported the outperformance of SMOTE. However, the learning time nearly doubled. In four candidate data sets, RS-I1, RS-I2, DT-I1, and RF-I1, RS-I2 outperforms the other three subsets. Therefore, we will use RS-I2 as our selected factor set.

This study adopted project content from the fundraising platforms Indiegogo and Kickstarter as a source of research data. The results obtained by using the above methods are shown in Table 16. From this table, we can see that factors X2, X3, X4, X9, and X12 are important factors for both the Indiegogo and Kickstarter crowdfunding platforms. Therefore, these five factors are listed as key factors for the success of the movie crowdfunding project.

Table 16. Summary of key factors for crowdfunding platforms.

Dataset (Subset) / Feature Set	Factor
Indiegogo (RS-I2) (SMOTE)	Role (X2), Cast (X3), Merchandise (X4), Traditional Advertising (X5), Social Media (X6), Funding (X7), Screen Features (X8), Sound effects (X9), Positive Sentiment (X10), Negative Sentiment (X11), Sentiment (X12)
Kickstarter (DT-K1)	Role (X2), Cast (X3), Merchandise (X4), Sound Effects (X9), Sentiment (positive–negative) (X12)

Based on the results, we provide some suggestions shown in Table 17 to fundraisers on crowdfunding platforms. Regarding "Role" and "Cast", which are the conventional crucial factors in successful movies, fundraisers have to mention more about the roles and characteristics of the movie characters, and should highlight the special actors and famous directors who can be mentioned more to attract the attention of investors/fans and, thus, increase the fundraising success rate. Another main source of benefit for a successful movie is "merchandise", so fundraisers should give more details of commemorative merchandise, clothing, movie soundtracks, and so on in the content of project description. Regarding "Sound effects", project content should mention more details of sound effects, such as classical, musical instruments, stereo effects, etc. Finally, the positive sentiment of project descriptions is very important for the success of movie projects, no matter whether Indiegogo or Kickstarter is being used. Therefore, fundraisers should keep project descriptions positive by using more positive words.

Compared to [7], which also focused on movie crowdfunding projects, this study only focus on project content. We can provide structured and concrete suggestions for fundraisers to write their proposal to enhance the success rate. Furthermore, compared to other works which studied the success rate of crowdfunding projects, such as [1–4], this study utilized social media reviews as experimental data without using questionnaires to collect data. The proposed method is more suitable for a big data environment and can obtain instant results for the voices of customers.

Table 17. Important key factors and suggestions.

	Key Factor	**Suggestion**
X2	Role	Fundraisers have to mention more about the roles and characteristics of the movie characters to attract investors' attention and increase the success rate of the project.
X3	Cast	For fundraisers, important or special actors and famous directors can be mentioned more to attract the attention of investors/fans and, thus, increase the fundraising success rate.
X4	Merchandise	Fundraisers have to mention more about the launch of merchandise, arouse the passion of collection for investors, and increase the success rate of fundraising.
X9	Sound effects	In the project content descriptions, fundraisers should mention what sound effects or tracks are used in the movie to obtain investors' attention and make them invest in the movie.
X12	Sentiment (positive–negative)	Fundraisers are suggested to use positive words in the project content rather than words with negative sentiments to increase the fundraising success rate.

6. Conclusions

To sum up, previous research only discussed project characteristics that affect the success rates of crowdfunding. Relevant research showed that the description of the project content is one of the key factors for the project success. What factors should be included in the project content, though, has not been discovered in most research. The goal of this study is to identify the key factors that contribute to a movie crowdfunding project's success. Twelve candidate factors were defined. Natural language processing and feature selection methods, including rough set theory, decision trees, and ReliefF are used to select optimal subsets of features by using real projects from the famous crowdfunding platforms Indiegogo and Kickstarter. Support vector machines are then used to assess the performance of selected factor subsets. Finally, five important key factors are identified, namely "Role", "Cast", "Merchandise", "Sound Effects", and "Sentiment (positive-negative)". Based on these 5 important key factors, relevant suggestions are made for future project sponsors to improve the success rate of crowdfunding projects.

Regarding the potential directions of future research, this study only concentrated on the three primary types of movie projects (comedy, narrative movie, and drama). In the future, the movie project can be expanded to other types. In addition, other fundraising platforms can be added to increase the number of research studies and improve the accuracy of the analysis. By choosing different feature selection methods, it may be possible to find more suitable feature selection methods in future research. Furthermore, future research needs to consider expanding the lexicon and making the research results more accurate. Concerning the limitations of this study, because we cannot find any study about movie crowdfunding projects, we only collected 12 candidate factors from the movie and crowdfunding literature. Therefore, we can only discover important factors from this candidate factor set. In the future, if more related works about success factors for movie crowdfunding movie projects are published, we can update the candidate factor set.

Author Contributions: Conceptualization, L.-S.C. and K.-F.Y.; methodology, Y.-R.L.; software, Y.-R.L.; validation, Y.-R.L. and L.-S.C.; formal analysis, Y.-R.L.; writing—original draft preparation, K.-F.Y. and Y.-R.L.; writing—review and editing, K.-F.Y. and L.-S.C.; visualization, K.-F.Y. and Y.-R.L.; supervision, L.-S.C.; project administration, L.-S.C.; funding acquisition, L.-S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by National Science and Technology Council, Taiwan (Grant No. MOST 111-2410-H-324-006).

Data Availability Statement: Data available on request.

Acknowledgments: Authors are grateful for the financial assistance provided by the National Science and Technology Council, Taiwan.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, X.; Liu, X.; Wang, X.; Zhao, H.; Zhang, W. Exploring the effects of social capital on crowdfunding performance: A holistic analysis from the empirical and predictive views. *Comput. Hum. Behav.* **2022**, *126*, 107011. [CrossRef]
2. Shneor, R.; Munim, Z.H. Reward crowdfunding contribution as planned behaviour: An extended framework. *J. Bus. Res.* **2019**, *103*, 56–70. [CrossRef]
3. Chandler, J.A.; Fan, G.; Payne, G.T. Working the crowd: Leveraging podcasts to enhance crowdfunding success. *Bus. Horiz.* **2022**, *65*, 79–88. [CrossRef]
4. Zhang, H.; Chen, W. Crowdfunding technological innovations: Interaction between consumer benefits and rewards. *Technovation* **2019**, *84–85*, 11–20. [CrossRef]
5. Belleflamme, P.; Lambert, T.; Schwienbacher, A. Crowdfunding: Tapping the right crowd. *J. Bus. Ventur.* **2014**, *29*, 585–609. [CrossRef]
6. Chang, J.R.; Chen, M.Y.; Chen, L.S.; Tseng, S.C. Why customers don't revisit in tourism and hospitality industry? *IEEE Access* **2019**, *7*, 146588–146606. [CrossRef]
7. Chen, M.Y.; Chang, J.R.; Chen, L.S.; Chuang, Y.J. Identifying the key success factors of movie projects in crowdfunding. *Multimed. Tools Appl.* **2022**, *81*, 27711–27736. [CrossRef]
8. Mina, F.I.; Baber, H. Crowdfunding model for financing movies and web series. *Int. J. Innov. Stud.* **2021**, *5*, 99–105.
9. Anglin, A.H.; Pidduck, R.J. Choose your words carefully: Harnessing the language of crowdfunding for success. *Bus. Horiz.* **2022**, *65*, 43–58. [CrossRef]
10. Lin, Y.; Boh, W.F. Informational cues or content? Examining project funding decisions by crowdfunders. *Inf. Manag.* **2021**, *58*, 103499. [CrossRef]
11. Chen, M.Y.; Chang, J.R.; Chen, L.S.; Shen, E.L. The key successful factors of video and mobile game crowdfunding projects using a lexicon-based feature selection approach. *J. Ambient Intell. Humaniz. Comput.* **2022**, *13*, 3083–3101. [CrossRef] [PubMed]
12. Schuckert, M.; Liu, X.; Law, R. A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently. *Int. J. Hosp. Manag.* **2015**, *48*, 143–149. [CrossRef]
13. Chen, W.K.; Chen, L.S.; Pan, Y.T. A text mining-based framework to discover the important factors in text reviews for predicting the views of live streaming. *Appl. Soft Comput.* **2021**, *111*, 107704. [CrossRef]
14. Verma, G.; Verma, H. Predicting Bollywood Movies Success Using Machine Learning Technique. In Proceedings of the 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 4–6 February 2019; p. 18620811.
15. Wang, Z.; Zhang, J.; Ji, S.; Meng, C.; Li, T.; Zheng, Y. Predicting and ranking box office revenue of movies based on big data. *Inf. Fusion* **2020**, *60*, 25–40. [CrossRef]
16. Kang, L.; Peng, F.; Anwar, S. All that glitters is not gold: Do movie quality and contents influence box-office revenues in China? *J. Policy Model.* **2022**, *44*, 492–510. [CrossRef]
17. Wei, L.; Yang, Y. An empirical investigation of director selection in movie preproduction: A two-sided matching approach. *Int. J. Res. Mark.* **2022**, *39*, 888–906. [CrossRef]
18. Zhang, Y.; Zhang, L. Movie Recommendation Algorithm Based on Sentiment Analysis and LDA. *Procedia Comput. Sci.* **2022**, *199*, 871–878. [CrossRef]
19. Moon, S.; Jalali, N.; Song, R. Green-lighting scripts in the movie pre-production stage: An application of consumption experience carryover theory. *J. Bus. Res.* **2022**, *140*, 332–345. [CrossRef]
20. Dreisbach, C.; Koleck, T.A.; Bourne, P.E.; Bakken, S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *Int. J. Med. Inform.* **2019**, *125*, 37–46. [CrossRef]
21. Thomaz, G.M.; Biz, A.A.; Bettoni, E.M.; Filho, L.M.; Buhalis, D. Content mining framework in social media: A FIFA world cup 2014 case analysis. *Inf. Manag.* **2017**, *54*, 786–801. [CrossRef]
22. Turban, E.; Aronson, J.E.; Liang, T.P.; Sharda, R. *Decision Support and Business Intelligence Systems*; Pearson Education Limited: London, UK, 2013.

23. Wang, W.; Zhu, K.; Wang, H.; Wu, Y.C.J. The Impact of Sentiment Orientations on Successful Crowdfunding Campaigns through Text Analytics. *IET Softw.* **2017**, *11*, 229–238. [CrossRef]

24. Du, Q.; Zhou, M.J.; Zhang, X.; Qiao, Z.; Wang, G.A.; Fan, W. Money Talks: A Predictive Model on Crowdfunding Success Using Project Description. In Proceedings of the Twenty-First Americas Conference on Information Systems, Fajardo, Puerto Rico, 13–15 August 2015.

25. Loureiro, S.M.C.; Guerreiro, J.; Ali, F. 20 years of research on virtual reality and augmented reality in tourism context: A text-mining approach. *Tour. Manag.* **2020**, *77*, 104028. [CrossRef]

26. Zhong, B.; Pan, X.; Love, P.E.D.; Sun, J.; Tao, C. Hazard analysis: A deep learning and text mining framework for accident prevention. *Adv. Eng. Inform.* **2020**, *46*, 101152. [CrossRef]

27. Alishahi, M.; Moghtadaiee, V.; Navidan, H. Add Noise to Remove Noise: Local Differential Privacy for Feature Selection. *Comput. Secur.* **2022**, *123*, 102934. [CrossRef]

28. Pawlak, Z. Rough sets and fuzzy sets. *Fuzzy Sets Syst.* **1985**, *17*, 99–102. [CrossRef]

29. Ho, I.L.; Shih, S.C.; Tsai, C.P.; Nagai, M. Decision Model Based on Grey System Theory and Rough Sets. *Int. J. Kansei Inf.* **2018**, *9*, 43–53.

30. Lei, L.; Chen, W.; Wu, B.; Chen, C.; Liu, W. A building energy consumption prediction model based on rough set theory and deep learning algorithms. *Energy Build.* **2021**, *240*, 110886. [CrossRef]

31. Su, C.T.; Chen, L.S.; Chiang, T.L. A neural network based information granulation approach to shorten the cellular phone test process. *Comput. Ind.* **2006**, *57*, 412–423. [CrossRef]

32. Albuquerque, L.G.; Roque, F.d.O.; Francisco, V.N.; Koroiva, R.; Buss, D.F.; Baptista, D.F.; Hepp, L.U.; Kuhlmann, M.L.; Sundar, S.; Covich, A.P.; et al. Large-scale prediction of tropical stream water quality using Rough Sets Theory. *Ecol. Inform.* **2021**, *61*, 101226. [CrossRef]

33. Pavitha, N.; Pungliya, V.; Raut, A.; Bhonsle, R.; Purohit, A.; Patel, A.; Shashidhar, R. Movie recommendation and sentiment analysis using machine learning. *Glob. Transit. Proc.* **2022**, *3*, 279–284. [CrossRef]

34. Khan, J. Weighted entropy and modified MDL for compression and denoising data in smart grid. *Int. J. Electr. Power Energy Syst.* **2021**, *133*, 107089. [CrossRef]

35. Yuan, Z.; Chen, H.; Xie, P.; Zhang, P.; Liu, J.; Li, T. Attribute reduction methods in fuzzy rough set theory: An overview, comparative experiments, and new directions. *Appl. Soft Comput.* **2021**, *107*, 107353. [CrossRef]

36. Rao, H.; Shi, X.; Rodrigue, A.K.; Feng, J.; Xia, Y.; Elhoseny, M.; Yuan, X.; Gu, L. Feature selection based on artificial bee colony and gradient boosting decision tree. *Appl. Soft Comput.* **2019**, *74*, 634–642. [CrossRef]

37. Shi, S.; Li, G.; Chen, H.; Liu, J.; Hu, Y.; Xing, L.; Hu, W. Refrigerant charge fault diagnosis in the VRF system using Bayesian artificial neural network combined with ReliefF filter. *Appl. Therm. Eng.* **2017**, *112*, 698–706. [CrossRef]

38. Jin, L.; Zeng, Q.; He, J.; Feng, Y.; Zhou, S.; Wu, Y. A ReliefF-SVM-based method for marking dopamine-based disease characteristics: A study on SWEDD and Parkinson's disease. *Behav. Brain Res.* **2019**, *356*, 400–407. [CrossRef]

39. Aslan, N.; Koca, G.O.; Kobat, M.A.; Sengul Dogan, S. Multi-classification deep CNN model for diagnosing COVID-19 using iterative neighborhood component analysis and iterative ReliefF feature selection techniques with X-ray images. *Chemom. Intell. Lab. Syst.* **2022**, *224*, 104539. [CrossRef]

40. Wen, X.; Xu, Z. Wind turbine fault diagnosis based on ReliefF-PCA and DNN. *Expert Syst. Appl.* **2021**, *178*, 115016. [CrossRef]

41. Kilicarslan, S.; Adem, K.; Celik, M. Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network. *Med. Hypotheses* **2020**, *137*, 109577. [CrossRef]

42. Souza, A.; Rojas, M.Z.; Yang, Y.; Lee, L.; Hoagland, L. Classifying cadmium contaminated leafy vegetables using hyperspectral imaging and machine learning. *Heliyon* **2022**, *8*, e12256. [CrossRef]

43. Zhang, B.; Li, Y.; Chai, Z. A novel random multi-subspace based ReliefF for feature selection. *Knowl.-Based Syst.* **2022**, *252*, 109400. [CrossRef]