Expert Systems with Applications 42 (2015) 3176–3193

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# Pre-production forecasting of movie revenues with a dynamic artificial neural network

M. Ghiassi *, David Lio, Brian Moon

*Professor of Information Systems Operations Management & Information Systems Santa Clara University 500 El Camino Real Santa Clara, Ca 95053-0388, USA*

A R T I C L E   I N F O

*Article history:*
Available online 20 November 2014

*Keywords:*
Movies
Box-office forecasting
Dynamic artificial neural networks
Artificial intelligence
Machine learning
Classification

A B S T R A C T

The production of a motion picture is an expensive, risky endeavor. During the five-year period from 2008 through 2012, approximately 90 films were released in the United States with production budgets in excess of $100 million. The majority of these films failed to recoup their production costs via gross domestic box office revenues. Existing decision support systems for pre-production analysis and green-lighting decisions lack sufficient accuracy to meaningfully assist decision makers in the film industry.

Established models focus primarily upon post-release and post-production forecasts. These models often rely upon opening weekend data and are reasonably accurate but only if data up until the moment of release is included. A forecast made immediately prior to the debut of a film, however, is of limited value to stakeholders because it can only influence late-stage adjustments to advertising or distribution strategies and little else.

In this paper we present the development of a model based upon a dynamic artificial neural network (DAN2) for the forecasting of movie revenues during the pre-production period. We first demonstrate the effectiveness of DAN2 and show that DAN2 improves box-office revenue forecasting accuracy by 32.8% over existing models. Subsequently, we offer an alternative modeling strategy by adding production budgets, pre-release advertising expenditures, runtime, and seasonality to the predictive variables. This alternative model produces excellent forecasting accuracy values of 94.1%.



## 1. Introduction & background

In this paper we present a forecasting model for the domestic (U.S.) box-office revenues of major motion pictures. To generate these forecasts we employ a dynamic artificial neural network model (DAN2) and use production, distribution, and advertising data as our primary inputs. Existing post-production and post-release forecasting models are currently capable of accurate projections. However, the low error rates of these models are achieved either with word-of-mouth data immediately prior to release or with post-release data from the opening weekend of a film. Unfortunately, forecasts at such a late stage are of little value to the studio or investor whose money has already been spent. Accordingly, our research is focused exclusively on the pre-production forecasting of potential box-office revenues.

The production of motion pictures is a risky business (De Vany & Walls, 2004); only between thirty and forty percent of films break even and just one out of ten films becomes profitable at the box-office (Hennig-Thurau, Houston, & Walsh, 2007; Vogel, 2001). Additionally, significant capital outflows are required well in advance of any anticipated revenues (i.e. ticket sales) and the product being produced is likely only to be viewed once in a theater by any particular person (Edwards, Buckmire, & Ortega-Gingrich, 2013).

The film industry is characterized by regular losses offset by infrequent blockbusters. In 2012, 68.8% of gross box-office revenues were generated by just 52 films, or 10% of all films released that year (Nash Information Services, LLC). During the five-year period from 2008 through 2012, more than 700 films were produced in the United States, with an average budget of nearly $42.5 million (Nash Information Services, LLC). Of these films, more than half failed to earn back their production budgets in domestic box-office revenues, with an average deficit of nearly $19 million dollars. Box-office performance increasingly depends upon a small number of blockbusters (Eliashberg, Elberse, & Leenders, 2006). This facet of the industry highlights the importance of decision

* Corresponding author at: Professor of Information Systems Operations Management & Information Systems Santa Clara University 500 El Camino Real Santa Clara, Ca 95053-0388, USA. Tel.: +1 408 554 4687; fax: +1 408 554 5157.

*E-mail addresses:* mghiassi@scu.edu (M. Ghiassi), david@davidlio.com (D. Lio), bmoon@scu.edu (B. Moon).

http://dx.doi.org/10.1016/j.eswa.2014.11.022
0957-4174/

support tools for stakeholders as their investments become spread across fewer films while the number of potential film projects available to them continues to increase.

Approximately 40% of box-office revenues for the average motion picture are accounted for by the first week of theatrical release (Einav, 2007). Accordingly, forecasts of total box-office revenues can be made with increasing accuracy in the weeks following initial release. Box-office revenue patterns are highly regular within several distinct types of distribution patterns (Jedidi, Krider, & Weinberg, 1998). Primary patterns are either a wide-release that relies upon intensive distribution and promotion or a platform-release that involves an incremental buildup of distribution intensity (Chen, Chen, & Weinberg, 2013; Sawhney & Eliashberg, 1996). To date, literature has not offered a viable method which successfully predicts these patterns without utilizing post-release or post-production data. Consequently, predicting box-office receipts without such data during the pre-production and pre-release periods remains a critical issue for the film industry (Sharda & Delen, 2006).

Our research focuses upon improving the accuracy of pre-production revenue forecasts for use by film investors and industry executives. Our goal is to develop forecasting models which are accurate, timely, and meaningful enough to be of substantive value to decision makers in the film industry. If accurate box-office revenue forecasts can be made before significant investment in development or production, a movie studio could save millions by avoiding a single flop. Due to the scale of investment and expense involved in modern motion pictures, even a marginal increase in the success rate of the "green-lighting" process would bring remarkable financial and reputational benefits to the studios and stakeholders involved (Eliashberg et al., 2006). For instance, if a forecast were to provide scenario analysis whereby a film could be identified as capable of earning between \$40–50 million with a PG-13 rating and a release on the Thanksgiving holiday weekend, or earning between \$30–40 million with an R rating and a release date two weeks later, the more profitable MPAA rating and production schedule could be pursued. Alternatively, if during the course of such analyses no sufficiently profitable forecast was identified, the project could be discarded altogether before significant investment or expenditure occurred. Such early insights would help to prevent outright failures at the box-office. Additionally, decision support during the project selection process helps studios to avoid the opportunity costs associated with tying up limited development resources on flops or even on mediocre projects. Such an improvement to resource allocation is especially pertinent to the film industry because even the largest studios release just a few dozen films each year. Because most widely-distributed films do no better than to break even financially, a single movie can often be the difference between millions of dollars of profits or significant losses for a studio during any given year (Vogel, 2001).

One group of industry stakeholders in particular who would benefit from improved box-office forecasts are rental contract negotiators. Because many content licensing deals are finalized prior to theatrical release, those seeking the rights to films for streaming services, rentals, or reselling could employ the decision support of accurate box-office revenue forecasts to better manage new content acquisition. Additional industry stakeholders strongly interested in pre-production revenue forecasts include studio executives and film investors, both of whom are looking to optimize resource allocation amongst competing projects and to maximize returns on investments. Top-tier actors, who are inundated with offers of various roles, would benefit from pre-production forecasting analysis to ensure that their limited time is devoted to projects with a high likelihood to succeed. Ultimately, anyone whose financial or reputational fortunes are tied to the revenue of a film would benefit significantly from accurate forecasts during the pre-production period.

### 1.1. The United States motion picture industry

At present, the U.S. film industry is strongly defined by its seven major studios. Collectively, they held just over 84% market share as of the end of 2013 (Table 1). There is a precipitous drop off in market share between 'the big seven' and the rest of their competitors, with the eighth largest studio, the Weinstein Company, having just over 4% market share. The operational purview of a studio generally encompasses the four functions of finance, production, distribution, and advertisement (Vogel, 2001). Theater chains and individual theater sites distinct from the studios themselves assume subsequent exhibition responsibilities (Eliashberg et al., 2006).

Despite an anti-trust court decision mandating the divorce of motion picture distributors from the business of exhibition (United States v. Paramount Pictures & 79 (U.S. Supreme Court, 1948) the studios continue to wield considerable influence over exhibitors. Ticket price variability, for instance, would greatly benefit movie exhibitors, whose revenues are generated in large part by concession sales. However, admission prices to low budget and to high budget films alike remain uniform (Orbach & Einav, 2007). While it might be argued that uniform pricing is generally favorable to the exhibitor, this is not universally true and at least some price variability, even if experimental, should be observable in a free marketplace – yet none occurs.

The aggregate financials of box-office ticket sales in the United States are considerable in scope, with over a billion tickets sold for each of the past eighteen years and total annual revenues of between \$10 billion and \$11 billion in each of the past five years. In addition to a recent uptick in domestic revenues, international revenues are also on the rise, having increased 32% over the past five years (MPAA, 2012).

The film industry is characterized by high volatility, even amongst projects within the same studio. Both wildly successful blockbusters and catastrophic flops occur with great regularity. In this market environment, the share of revenues captured by blockbuster movies continues to rise (Eliashberg et al., 2006). The average production cost of a film has increased from under \$30 million in the early 1990s to almost \$65 million in the early 2000s (Eliashberg et al., 2006). Advertising budgets have jumped from roughly \$10 million to nearly \$35 million over the same period (MPAA, 2004). In the summer of 2004 alone, Hollywood studios released more than two dozen films with production budgets in excess of \$100 million (Marr & Orwall, 2004).

Between 2000 and 2010, 75 movies each had deficits greater than \$50 million between their production budgets and their gross domestic box-office revenues. During the same period, of the 3453 films for which budgetary and domestic box-office data were available, 1111 (31%) failed to break even domestically, with an average loss of \$16.5 million (Table 2).

Over the same time period (2000–2010), as production and marketing costs increased so too did the level of financial risk facing equity investors (Desai, Loeb, & Veblen, 2002). As a consequence, studios increasingly have begun to capitalize on the brand equity of established franchises, resulting in a shift towards the production of sequels and trilogies as safer alternatives to untested ventures (Eliashberg et al., 2006).

### 1.2. Box-office revenues

Box-office revenues are not representative of all film-related revenue streams. For instance, many of the people who see a given film will not do so in a theater. Of non-theatrical viewers, some

**Table 1**
United States motion picture industry (2013).

| Rank | Company (studio) | # Movies | Total box-office | Average box-office | Market share (%) |
|---|---|---|---|---|---|
| 1 | Time Warner (Warner Bros.) | 29 | $ 1,861,194,799 | $ 64,179,131 | 17.08 |
| 2 | Walt Disney | 18 | 1,721,354,677 | 95,630,815 | 15.79 |
| 3 | GE/Comcast (Universal) | 19 | 1,415,663,293 | 74,508,594 | 12.99 |
| 4 | Sony Pictures | 20 | 1,149,187,808 | 57,459,390 | 10.54 |
| 5 | News Corp. (20th Century Fox) | 20 | 1,069,359,977 | 53,467,999 | 9.81 |
| 6 | Lionsgate | 20 | 1,017,528,833 | 50,876,442 | 9.34 |
| 7 | Viacom (Paramount Pictures) | 16 | 974,735,713 | 60,920,982 | 8.94 |
| | Total/Average | 142 | $ 9,209,025,100 | $ 64,852,289 | 84.49 |

**Table 2**
Aggregate box-office statistics (2000–2010).

| Total films released | 2079 Films |
|---|---|
| Average production budget ($ USD) | $ 35,737,733 |
| Average US box-office revenue ($ USD) | $ 44,213,244 |
| Average worldwide box-office revenue ($ USD) | $ 111,986,081 |
| **Gross domestic box-office receipts > production budget** | **968 films** |
| Average profit ($ USD) | $37,235,560 |
| **Gross domestic box-office receipts < production budget** | **1,111 films** |
| Average domestic loss ($ USD) | ($16,582,750) |
| **Gross global box-office receipts < production budget** | **445 films** |
| Average global loss ($ USD) | ($12,839,501) |

percentage will purchase or rent the content after its release to home video formats (DVD, Blu-Ray, cable video on demand, digital download, etc.). Others might encounter the content as part of a subscription service (e.g. Netflix, Hulu, cable television, etc.). Even those who pirate films might sometimes purchase film-affiliated merchandise. Accordingly, box-office receipts are but one of many revenue streams and a film with box-office revenues less than its production budget may nonetheless become profitable through other avenues.

The theatrical performance of a movie in the United States is a critical driver of success in subsequent release windows and channels. Films are first distributed to the market that generates the highest revenue over the shortest time. Releases then cascade downward in order of revenue contribution. Historically, this has meant that an initial, domestic, theatrical release was followed by international releases, pay-cable programming, home video, network television, and eventual local television syndication (Elberse & Eliashberg, 2003; Vogel, 2001). Regardless of the mechanism of consumption, the fundamental proposition of creating, developing and delivering an experiential product such as a motion picture remains the same.

In almost all instances, the box-office revenues of a film are strongly and positively correlated with other primary revenue streams (Elberse & Eliashberg, 2003; Litman & Ahn, 1998; Terry, Cooley, & Zachary, 2009). Therefore, when assessing the profitability of a motion picture, accurate forecasting of box-office revenue becomes the essential step in overall financial calculations for such an investment. We will show DAN2 to be an effective tool for forecasting box-office revenues for motion pictures in the United States. Our initial work follows the methodology of Sharda and Delen (2006, 2010) who converted this forecasting problem into a classification problem by categorizing films into one of nine classes, ranging from 'flop' to 'blockbuster'. Our replication of their experiments demonstrates the viability of DAN2's capacity to function as a more accurate classifier in this domain.

In subsequent experiments we demonstrate how alternative models offer further improvements to forecasting accuracy. Foundational to this research are the datasets we have constructed in part by harvesting a range of publicly available film data. Additionally, we have purchased data from industry sources and obtained data from other researchers. Consequently, our datasets are more comprehensive and longitudinal in nature than those previously cited in the research literature.

In all of our models, our dependent variable is the total domestic revenue of ticket sales; this excludes auxiliary revenues such as video rentals, international market revenues, merchandise and soundtrack sales, etc. We will consider revenues generated from ticket sales only during the full run of an initial theatrical release. Due to the relative scarcity of re-released films and a resultant dearth of available data we have excluded films with multiple theatrical releases from our experiments. In a similar vein, we choose also to exclude 3D and IMAX releases from our analyses due to their specific theatrical requirements and a similar lack of available data.

The remainder of this paper is organized accordingly. In Section 2 we present an overview of the current literature on the topic of box-office revenue forecasting, examine modeling approaches from a variety of disciplines, and review the current state of the art. Subsequently, we discuss artificial neural networks and justify their application to this problem along with the implementation performed by Sharda and Delen (2006, 2010). In Sections 3 and 4 we discuss our replication of this experiment and the use of DAN2 as an alternative model. We next consider the impact of introducing new variables in general and advertising expenditures in particular, upon forecast accuracy. Finally, in Sections 5 and 6, we review our results, consider their implications, and suggest future research opportunities.

## 2. Literature review

Revenue forecasting is a broadly studied domain within the film industry. Significant research has been conducted by econometricians, marketing professionals, and operations researchers, as well as by word-of-mouth experts and neural network scientists. In our review of extant literature we consider both forecasting efforts and variate analyses. Based upon the temporal availability of inputs to various models, we also consider as significant the time frame during which a given forecasting approach is viable: post-release, post-production (yet still pre-release), and pre-production. Several domain overviews and meta-analyses of previous research offer syntheses of ongoing research trends and established theories (Eliashberg, Weinberg, & Hui, 2008; Eliashberg et al., 2006; Karniouchina, 2008; McKenzie, 2012).

### 2.1. Econometric-based studies

Econometric models identifying variables of predictive value for box-office revenue forecasting include regression analyses of movie characteristics and explorations of endogeneity between variables (Elberse & Eliashberg, 2003; Hennig-Thurau et al., 2007; Litman, 1983; Neelamegham & Chintagunta, 1999; Prag & Casavant, 1994).

A focal area of econometric research in this domain has been the analysis of the performances of sequels relative to their parent films as well as relative to contemporaneously released non-sequel films (Basuroy & Chatterjee, 2008; Dhar, Sun, & Weinberg, 2012; Hennig-Thurau, Houston, & Heitjans, 2009; Karniouchina, 2008; Moon, Bergey, & Iacobucci, 2010).

Research into the effects of movie stars upon the performance of motion pictures at the box-office is one of the most well researched areas within the film industry (Ainslie, Drèze, & Zufryden, 2005; Basuroy, Chatterjee, & Ravid, 2003; Canterbery & Marvasti, 2001; De Vany & Walls, 2004; Elberse, 2007; Elberse & Eliashberg, 2003; Hennig-Thurau et al., 2007; Litman, 1983; Litman & Ahn, 1998; Litman & Kohl, 1989; Liu, 2006; Neelamegham & Chintagunta, 1999; Nelson & Glotfelty, 2012; Ravid, 1999; Ravid & Basuroy, 2004; Sawhney & Eliashberg, 1996; Sochay, 1994; Walls, 2005).

One class of econometric research in this domain is behavioral modeling which focuses upon the process of film selection from contemporaneously competing films and alternative entertainment venues (De Silva, 1998; Eliashberg, Jonker, Sawhney, & Wierenga, 2000; Sawhney & Eliashberg, 1996; Zufryden, 1996). Relatedly, extant research has identified significant seasonal patterns within the film industry that both influence and are influenced by the release patterns of the major studios (Einav, 2007; Einav, 2010; Einav & Ravid, 2009; Elberse & Eliashberg, 2003; Gutierrez-Navratil, Fernandez-Blanco, Orea, & Prieto-Rodrigues, 2012; Krider & Weinberg, 1998).

MOVIEMOD, a post-production, pre-release market evaluation model driven by Markov-chains, has achieved forecasts with error rates as low as 10% across a sample of 15 films (Eliashberg et al., 2000). Limited to analyzing the penetration and performance of films in non-US markets, this research employs trial viewings and subsequent surveys among target populations via consumer-clinics. This both tests existing awareness and also introduces the concept of the film to those who were unfamiliar. As a decision support system for the distribution of motion pictures in ancillary markets subsequent to initial release, MOVIEMOD appears to have been an effective predictive system in the markets in which it has been tested.

### 2.2. Marketing-based studies

Much of the marketing research conducted in the domain of box-office revenue forecasting involves the examination of the role of online reviews (Dellarocas, Farag, & Zhang, 2005; Dellarocas, Zhang, & Awad, 2007; Duan, Gu, & Whinston, 2008a; Duan, Gu, & Whinston, 2008b). In particular, examinations of the word-of-mouth features found within the Hollywood Stock Exchange (HSX) have yielded highly accurate post-production and post-release forecasts of box-office revenues (Doshi, 2010; Elberse & Anand, 2007; Elberse & Eliashberg, 2003; Foutz & Jank, 2007; McKenzie, 2013; Spann & Skiera, 2003).

Marketing research has included examinations of market share based forecasting models (Ainslie et al., 2005). Sequel analysis and examinations of the relative performance of sequels also have been discussed in the marketing literature (Dhar et al., 2012). For our analysis of the predictive power of advertising expenditures for box-office revenues we expanded upon extant research in this domain (Einav, 2007; Elberse & Anand, 2007; Gopinath, Chintagunta, & Venkataraman, 2013; Hanssens, Gilbert-Rolfe, Merchant, & Moroian, 2003; Hennig-Thurau et al., 2007; MPAA, 2007; Prag & Casavant, 1994; Rennhoff & Wilbur, 2010) and have used various industry studies and polls that were publicly available (Fetto, 2010; MPAA, 2004; MPAA, 2007; MPAA, 2012). Overviews of marketing literature as it pertains to the motion picture industry have also been published in recent years (Eliashberg et al., 2006; Eliashberg et al., 2008).

### 2.3. Operations-based studies

In contrast with aggregate forecasts, theater-level box-office revenue forecasts have been conducted for the purpose of optimal theater selection at the time of theatrical release (Somlo, Rajaram, & Ahmadi, 2011). Demographic, geographic, and theater-specific variables (such as proximity to other theaters) allowed for a distribution planning support system that claims the potential to increase average distributor profits by 12% or $1.8 million per film. Demographic information about movie goers, such as the data employed by Somlo et al. (2011) could be employed in future research on aggregate forecasts to determine more accurate measures of competition amongst contemporaneously released films. Econometric research on film release regions has also sought to model geography-based effects on movie goers and exhibitor behavior (Edwards et al., 2013).

### 2.4. Word-of-mouth & social media studies

Although word-of-mouth (WOM) was the most widely discussed variable pertaining to box-office revenue, our own experiments do not utilize data of this type so as to preserve our capacity for pre-production forecasting. Two primary categories of WOM data exist for films: reviews by movie critics and consumer-generated WOM. With WOM being largely a function of advertising, press, and general "buzz", little if any WOM data can be gathered before a film has been advertised, limiting the decision support value of such WOM based forecasts. Nevertheless, forecasting models employing WOM data are very accurate during the post-production and post-release periods.

Although positive critical reviews are correlated with higher box-office revenues (Terry, Butler, & De'Armond, 2005), this relationship is far weaker than the correlation of negative critical reviews with lower box-office revenues (Basuroy et al., 2003). A similar pattern is observable in the relationship between online consumer reviews of books and relative sales, whereby the negative impact of a low rating is greater than the positive impact of an equivalently high rating (Chevalier & Mayzlin, 2006).

Correlations have been consistently identified between consumer-generated WOM volume and box-office revenues (Asur & Huberman, 2010a; Asur & Huberman, 2010b; Dellarocas et al., 2005; Dellarocas et al., 2007; Duan et al., 2008a; Joshi, Das, Gimpel, & Smith, 2010; Liu, 2006; Liu et al., 2010). WOM valence (sentiment) is found by some researchers to be of little or no predictive value (Duan et al., 2008a; Duan et al., 2008b; Joshi et al., 2010; Liu et al., 2010). Other researchers, however, find that consumer-generated WOM valence may be a significant or contributory predictor of box-office performance (Asur & Huberman, 2010b; Dellarocas et al., 2005; Dellarocas et al., 2007; Rui, Liu, & Whinston, 2010). It has also been observed that WOM valence can influence WOM volume, having an indirect effect on box-office revenues (Duan et al., 2008a). For example, contentious or controversial films with a strong valence, whether positive, negative, or both, are likely to spur further discourse. Before a film has been released, since no one but a few critics will have actually seen the film, most WOM is bound to be speculative. Accordingly, consumers are not generally influenced by the persuasive effects of online WOM but they are affected in terms of awareness (Duan et al., 2008a).

Recent research offers some nuanced insight into the potential dynamics of professional reviews. Derrick, Williams, and Scott (2014) have found that for reviews of movie critics made available before the release of a film, volume alone is positively correlated with opening week box-office revenues. However, for second and subsequent weeks, both the volume and valence of professional reviews are positively correlated with weekly box-office revenues. Related research suggests that while WOM volume and WOM valence appear to be significant factors for box-office outcomes of American films in the U.S. domestic market, only WOM volume appears to be a significant factor for American films released in international markets (Kim, Park, & Park, 2013). However, for international films exhibited in their own native markets, both WOM volume and valence appear to be of predictive value as evidenced by studies of the influence of social network services within the Korean film market (Kim, Hong, & Koo, 2013). Additionally, while both expert reviews and online WOM are of predictive value for American films released in the U.S. domestic market, only online WOM appears to be of predictive value for American films released in international markets (Kim et al., 2013).

In addition to critical reviews, consumer ratings, and tweets, other online WOM that have been studied in this domain include the volume of traffic handled by a film's official website (Zufryden, 2000), film references on blogs (Sadikov, Parameswaran, & Venetis, 2009), the volume of internet search queries pertaining to a film (Hand & Judge, 2012; Kulkarni, Kannan, & Moe, 2012; Panaligan & Chen, 2013), news article analysis (Zhang & Skiena, 2009), text analysis of critical reviews (Joshi et al., 2010), sentiment analysis of YouTube trailer comments (Apala et al., 2013), and the amount of activity in the discussion section of a film's Wikipedia article (Mestyán, Yasseri, & Kertész, 2012a, 2012b) or the web traffic to a film's Wikipedia article (De Silva & Compton, 2014), all measured in the days and weeks preceding theatrical release.

At present, the most effective WOM-based forecasting models utilize data from the Twitter messaging service. The rate at which tweets concerning a particular film are generated ('tweet velocity') has been found to be of value for determining box-office revenues (Asur & Huberman, 2010b). 'Tweet valence' (sentiment) was found only to improve box-office revenue predictions during the post-release period. Independent research has also confirmed the value of such twitter data to this domain (Guardia-Sebaoun, Rafrafi, Guigue, & Gallinari 2013; Rui et al., 2010). A time series of tweet velocity measurements for the 7 days prior to release, when combined with the number of theaters at which a film is to be released, was able to achieve Adjusted Mean Absolute Percentage/Relative Error (AMAPE) of less than 1% in a regression model (Asur & Huberman, 2010b). These results were obtained with a dataset of 24 movies released during a four month period between November 2009 and February 2010; this methodology currently represents the state-of-the-art in post-production box-office revenue forecasting. While Twitter data appears to be predictive of box-office performance, additional research has shown that aggregate tweet valence might not be directly correlated with eventual online ratings (Wong, Sen, & Chiang, 2012).

The Hollywood Stock Exchange (HSX) is a virtual stock exchange that allows for the trading of 'shares' of films, actors, filmmakers, and other similar entities. The exchange is treated as a game and new players are given free virtual funds to play-invest with. HSX's current owner, Cantor Fitzgerald & Company, manages real film gambling interests in Europe and likely uses the HSX to assist in the management of these operations.

Virtual stock markets are utilized for forecasting in various other domains, including sports and political elections. Algorithmic and statistical approaches capitalizing upon the wisdom of the crowd found in virtual stock markets have been demonstrated to outperform the predictions of singular, expert, human opinion (Doshi, 2010; Spann & Skiera, 2003). Statistical analysis of HSX data demonstrates the value of prediction-markets for forecasting box-office revenues, despite the tendency for small-earning films to be over-predicted and large-earning films to be under-predicted (McKenzie, 2013). Analytical methods traditionally employed by those working within real stock markets, such as functional shape analysis, have been applied to the virtual stock market of the HSX (Foutz & Jank, 2007). In this model, graphs are created by plotting the price of a movie share over time as the release date approaches. Shape analyses are then conducted on these patterns, from which categories are established for the future classification of new films. The analyses of these patterns include calculations of price velocity, trading volume, and other metrics commonly employed in traditional stock markets. Box-office revenue forecasts have been demonstrated with MAPE rates of less than 8.5% by taking advantage of all the information generated up until the moment of theatrical release (Foutz & Jank, 2007). Error rates (MAPE) of approximately 36% were recorded for forecasts performed 40 weeks prior to the release of a film, with these error rates declining over the intervening period.

### 2.5. Forecasting studies

Box-office revenue forecasts can be distinguished by the availability timing of their requisite input data. Alternatively, forecasting efforts can also be categorized according to the methodology employed. Bayesian models, regression-based models, and artificial neural networks (ANNs) are the primary types of approaches demonstrated in extant literature on box-office revenue forecasting.

Most forecasts are conducted with data available during the post-production, pre-release period (Ainslie et al., 2005; Asur & Huberman, 2010a; Asur & Huberman, 2010b; De Silva, 1998; Dellarocas et al., 2007; Eliashberg, Hui, & Zhang, 2007; Eliashberg et al., 2000; Kulkarni et al., 2012; Litman, 1983; Litman & Kohl, 1989; Marshall, Dockendorff, & Ibanez, 2013; Neelamegham & Chintagunta, 1999; Parimi & Caragea, 2013; Sawhney & Eliashberg, 1996; Sochay, 1994; Zufryden, 1996). Post-release forecasting models are conducted after initial theatrical release, when at least the first week of receipts are available (Dellarocas et al., 2005; Derrick et al., 2014; Neelamegham & Chintagunta, 1999; Ravid, 1999; Sawhney & Eliashberg, 1996).

Some of the most successful research into box-office revenue forecasting with Bayesian methods has been conducted by Neelamegham and Chintagunta (1999), whose pre-release model significantly improved upon the forecasts of Sawhney and Eliashberg (1996). Mean absolute percentage error (MAPE) rate of domestic pre-release box-office forecasts across 10 unreleased movies was improved by Neelamegham and Chintagunta from Sawhney and Eliashberg's 71% to their 45%.

Similar, more recent results were also obtained by Marshall et al. (2013) who incorporated Sawhney and Eliashberg's methods into their own model based upon longstanding econometric modeling techniques for new product adoption (Bass 1969). Although limited to a subset of films released in Chile between 2001 and 2003, pre-release forecasts of attendance were reported by the authors with error rates of 55.65%. These error rates declined over time as post-release information (weekly attendance) was incorporated into their model. This resulted error rates of 5.63% by the sixth week and 2.72% by the twelfth week.

Textual analysis is another avenue of movie revenue forecasting study that has been pursued by Eliashberg et al. (2007), Eliashberg, Hui, and Zhang (2014). Initial experimentation, based solely upon the textual analysis of four 20-page film synopses (or 'spoilers') for 281 films released between 2001 and 2004, achieved a classification rate of just under 62% for the categories of 'below average', 'average', and 'above average' return on investment (revenue/budget) (2007). Subsequent experimentation based upon the textual analysis of the full scripts for 300 films released between 1995 and 2010 produced box-office revenue forecasts with mean squared error of approximately 40% (2014). Script analysis for box-office revenue forecasting holds great potential for improving pre-production forecast accuracy because it relies solely upon data available during the pre-production period (i.e. the screenplay itself).

Artificial neural network (ANN) models have been demonstrated to be an effective forecasting method in a variety of domains, including retail sales (Alon, Qi, & Sadowski, 2001). For box-office revenue forecasting in particular, the efficacy of neural network based models has been demonstrated, as has their superiority over competing methods, including logistic regression, support vector machines, and regression-based models (Delen & Sharda, 2010; Sharda & Delen, 2006; Zhang, Luo, & Yang, 2009). ANN-based models are most effective when the underlying relations are nonlinear. In every iteration ANN models try to solve the resulting nonlinear optimization and these can never be proven to be globally optimal. Using ANN to solve regression problems can result in suboptimal solutions whereas regression guarantees optimality. The most accurate, publicly available, box-office revenue forecasting results have been published by Delen and Sharda (2010) with an ANN-based model. Therefore, we too have chosen an ANN-based modeling approach to the box-office revenue forecasting problem. We next briefly present their most recent experiments for use as a baseline to measure the effectiveness of our model.

### 2.6. ANN-based forecasting models

The variables employed by Sharda and Delen in their analysis are as follows: MPAA rating, competition, star value, genre, special effects, sequel, and number of screens. MPAA ratings considered were: G, PG, PG-13, R, and not-rated. 'Competition', 'star value', and 'special effects' data are constrained to low, medium, and high values. The 'sequel' variable is a binary flag indicating whether or not a given film is a sequel. 'Number of screens' is an integer value representing the number of screens on which a movie is shown at the time of theatrical release.

#### 2.6.1. MPAA rating

The Motion Picture Association of America (MPAA) rating system has been in place for over 50 years. While it has evolved over time, its ratings and their definitions have remained relatively static over the past few decades (the period of our analysis). Due to legal restrictions on viewership by age, the MPAA rating of a film is a primary indicator of potential audience size and age demographics.

More restrictive MPAA ratings (R, for example) are negatively correlated with box-office revenues, likely due to the 'restricted' audience size (Terry, Butler, & De'Armond, 2005). MPAA ratings in general have been identified as significant predictors of box-office revenues in regression analysis (Basuroy & Chatterjee, 2008). While only six distinct MPAA ratings exist (G, PG, PG-13, R, NC-17, and Unrated), the justifications behind each particular rating decision might hold additional predictive power for box-office revenue forecasting. For instance, films with high levels of violence and violent films that also contain sexual content earn higher than average revenues, although they are not necessarily more profitable (Ravid & Basuroy, 2004). Accordingly, more granular analyses of rating decisions such as the occurrence and levels of controversial elements such as sex, violence, and drug use may be worth pursuing in future box-office forecasting research.

#### 2.6.2. Competition

A film, like any other good, competes in the marketplace for a specific and finite pool of consumer dollars. This competition occurs most strongly between films with similar release timings, MPAA ratings, and genres. Two PG-13 rated military action films released on the same holiday weekend are obviously in direct competition with one another. The box-office revenues of films that share both genre and release date tend to be negatively affected by such contemporaneous release scheduling (Ainslie et al., 2005; Gutierrez-Navratil et al., 2012). This effect is also observable for films with differing genres but identical MPAA ratings that are released contemporaneously. However, this sales loss becomes less severe in weeks subsequent to initial release (Ainslie et al., 2005).

As studios seek to capture maximum market share, especially during times of peak admission, they are often weighing early-release strategies against delayed-release strategies to avoid direct competition from the strongest films of other studios (Krider & Weinberg, 1998). Although publicly changing a previously announced release date negatively impacts the stock price of the studio (or holding company) responsible for the film, changes to release dates do not appear to be correlated with subsequent box-office revenues (Einav & Ravid, 2009). Despite this lack of an observable penalty to revenues for rescheduling release dates, studios regularly over-cluster their release dates, especially on holiday weekends (Einav, 2010). One explanation for this ostensibly non-optimal behavior might be attributable to the star-centric model of the industry, whereby studios would incur reputational penalties for not cultivating and reinforcing star power during times of peak attendance.

#### 2.6.3. Star value

Although widely discussed in the literature, a consensus has not been reached on the role and significance of star power in predicting box-office revenues. Star power is almost always highly correlated with production budget (Basuroy et al., 2003; Hennig-Thurau et al., 2007). Accordingly, the presence of a top-tier actor or actress generally can be inferred from a large production budget. Almost all films with production budgets in the hundreds of millions of dollars will employ at least one actor or actress with a high star-power value.

As it is employed in the papers surveyed herein, this star-power value is generally calculated by comparing the averages of the recent box-office revenues of films in which a star is credited with the averages of other stars. This metric is often limited to recent years and to films in which a star is prominently credited to avoid noise in the data from forgotten roles, cameos, and off-camera work. Other measures of star power are drawn from industry sources such as IMDB.com's 'StarMeter' rankings.

Due to a divergence in the methodologies employed when calculating star power, some of the research done on this topic is not necessarily directly comparable. However, in aggregate, a meta-analysis of the research reveals that while star power positively influences revenues, the relationship between star power and profitability is either minimal or negative.

Nine studies found a positive relationship between star power and opening or total revenues (Ainslie et al., 2005; Basuroy et al., 2003; Canterbery & Marvasti, 2001; Elberse, 2007; Elberse & Eliashberg, 2003; Neelamegham & Chintagunta, 1999; Nelson & Glotfelty, 2012; Sawhney & Eliashberg, 1996; Sochay, 1994). Six papers found either no relationship or a negative relationship between star power and profitability (Hennig-Thurau et al., 2007; Litman, 1983; Litman & Ahn, 1998; Liu, 2006; Ravid, 1999; Ravid & Basuroy, 2004). Additional research has observed that while star power may not affect opening week revenues, the star power of a given film is generally positively correlated with the revenues of second and subsequent weeks (Derrick et al., 2014).

Confirming the 'curse of the superstar' (De Vany & Walls, 2004), if a star is paid the expected increase in revenue associated with his or her performance then the movie will almost always lose money (Walls, 2005). The effects of star power upon revenue do not appear to extend to the level of celebrity possessed by the director (Nelson & Glotfelty, 2012). The effects of star power upon revenue do not appear to extend to the level of celebrity possessed by the director for American films in domestic markets (Nelson & Glotfelty, 2012). However, the impact of director star power may be a regional phenomenon, as it has been observed to be of predictive value for Korean films in their domestic markets (Song & Han, 2013).

Although researchers disagree on the precise impact of star power and revenues, both the larger budget required to obtain a high degree of star power as well as the star power itself can act as insurance policies by moderating the negative effects of potentially unfavorable critical reviews (Ravid & Basuroy, 2004). Additionally, films with both large budgets and high star power facilitate economies of scale for the highly vertically integrated studios, which depend in part upon large-scale productions to keep all of the various components of their expansive businesses engaged (Canterbery & Marvasti, 2001).

The relationship between star power and success has been hypothesized to be non-linear in nature (Hennig-Thurau et al., 2007). Accordingly, an artificial neural network is an appropriate tool for this problem domain as ANNs generally handle non-linearity more effectively than do competing methods such as linear and nonlinear regression.

### 2.6.4. Genre

Due to the complexity and depth of nearly any story that is told over the course of an hour or two, genre classifications for films are not mutually exclusive. The film *Men in Black*, for example, is as clearly comedic as it is a work of science fiction. Perhaps as a consequence of the complex, often plural nature of this genre variable, most research has either been inconclusive with regard to the impact of genre on box-office revenues (Elberse & Eliashberg, 2003) or has found that only some genres have significant effects on box-office revenues (Ainslie et al., 2005). Science-fiction and comedy are most commonly found to be positively associated with significant effects upon box-office performance (Litman, 1983; Liu, 2006; Moon et al., 2010; Prag & Casavant, 1994; Sochay, 1994; Terry, Butler, & De'Armond, 2005).

In the context of box-office forecasting, the genre of a film is an important attribute in determining prospective audience demographics. In conjunction with release timing and MPAA rating, the genre of a film can be used to help determine the number and nature of potential movie-goers. For example, a G-rated comedy released on a holiday weekend competes for a markedly different audience than an R-rated horror film released the following weekend.

### 2.6.5. Special effects

The level of special effects found in a film was a variable considered by Sharda & Delen on a scale of low/medium/high (Delen & Sharda, 2010; Sharda & Delen, 2006). Sawhney and Eliashberg (1996) were the only other researchers to include this variable; they did so as a binary flag indicating the presence or absence of major special technical effects.

### 2.6.6. Sequel

A film that is a sequel (or prequel, or part of a trilogy, etc.) is markedly different from other films in that a sequel, by its very definition, has at least one sibling data point. Sequels are generally based upon movies with better than average box-office revenues (Dhar et al., 2012). One notable advantage of sequels outside of the box-office are higher than average home video sales due to the brand awareness generated by parent films (Hennig-Thurau et al., 2009).

An analysis of ratings from Rotten Tomatoes and Yahoo Movies has found that, in general, sequels tend to reap more revenues but receive lower ratings than their original counterparts (Moon et al., 2010). Although sequels tend to outperform their contemporaneous non-sequel counterparts in terms of box-office revenues, attendance, and profitability, they do not tend to outperform preceding films within the same franchise (Basuroy & Chatterjee, 2008; Dhar et al., 2012). Due to the larger than normal audiences drawn by sequels, they tend to be shown on more screens than non-sequel films (Dhar et al., 2012). The advantage that sequels of prior films have over singular titles is correlated with the time between a sequel's release date and the release date of its parent film. This effect diminishes over time as mindshare fades and the brand of the franchise becomes increasingly difficult to leverage (Basuroy & Chatterjee, 2008). In other words, sequels tend to out-perform non-sequels due to both the pre-existing cultural familiarity they engender and the proven success of their precursors (Hennig-Thurau et al., 2007).

### 2.6.7. Screen count

The number of screens on which a film is to be shown at the time of its theatrical release is positively correlated with revenues (De Vany & Walls, 2004; Hennig-Thurau et al., 2007; Neelamegham & Chintagunta, 1999; Sochay, 1994). In the period after initial release, the number of screens on which a film is shown continues to be positively correlated with box-office revenues but diminishing effects can be observed once market saturation has been achieved (Basuroy & Chatterjee, 2008). In 2012, there were 39,918 active screens in the United States, about 3% more than there were in 2008 (MPAA, 2012). Of these screens, 7647 (19%) were located in 1–7 screen venues, while the remaining 32,271 (81%) screens were located in venues with eight or more screens (MPAA, 2012).

In addition to the initial number of screens, some researchers have examined metrics such as the peak number of screens (i.e. the greatest number of screens on which a film was shown throughout the course of its run in theaters). In most cases, the initial number of screens is also the peak number of screens, which we would expect given the typical distribution process of an escalating advertising campaign culminating in a nation-wide, weekend release. While a variable such as peak number of screens may be useful for the identification of 'sleepers' (i.e. films with a modest debut that end up with excellent box-office performances), such a figure is of less use to pre-release forecasting, which is by its very nature limited to considering only what can be known about a film before initial theatrical release.

### 2.6.8. Variable set construction

The movie metrics employed by Sharda and Delen were transformed from 7 variables into 26 data points for input into their ANN model (Table 3).

In Sharda and Delen's experiment a record has a column for each one of these 26 possible values. Possible values for all

**Table 3**
Sharda & Delen variables & ANN inputs.

| Variable | Possible Values |
|---|---|
| MPAA rating | G, PG, PG-13, R, NR |
| Competition | High, medium, low |
| Star value | High, medium, low |
| Genre | Sci-Fi, epic drama, modern drama, thriller, horror, comedy, cartoon, action, documentary |
| Special effects | High, medium, low |
| Sequel | Yes, no |
| Number of screens | A positive integer between 1 and 3876 |

**Table 4**
Sharda & Delen: classification thresholds.

| Class | Revenue range (in $ millions) |
|---|---|
| (Blockbuster) A | 200+ |
| B | 150–200 |
| C | 100–150 |
| D | 65–100 |
| E | 40–65 |
| F | 20–40 |
| G | 10–20 |
| H | 1–10 |
| (flop) I | <1 |

variables are binary flags (i.e. 0 or 1) except for the number of screens, which is included as the sole continuous variable. For example, a film with a high degree of star value that is also a sequel would have, among its other attributes, the following: '1' in the 'star value: high' column, '0' in both the 'star value: medium' and 'star value: low' columns, and '1' in the 'sequel' column. Accordingly, the single variable 'star value' is expanded into three possible values. This is the process whereby 7 initial variables become 26 possible values. Since ANNs are capable of processing both continuous and discrete values as inputs, benchmark experiments include continuous values for the number of screens variable because they result in better prediction results than did discretized values (Sharda & Delen, 2006). While values for MPAA rating, competition, star value, special effects, and sequel are mutually exclusive, a film can belong to more than one genre.

*2.6.9. Model construction*

In their ANN based approach to box-office revenue forecasting, Sharda and Delen have converted the problem of revenue forecasting from a point-estimate (i.e. a singular figure) into a classification problem (i.e. a revenue interval). In this way, movies are classified into one of nine classes from 'flop' to 'blockbuster' (Table 4). This clustering of films allows for an ANN based model to be trained to recognize elements and combinations of elements that are of predictive value from similarly performing films.

*2.6.10. Accuracy metrics*

The model uses average percent hit rate (APHR) as the accuracy metric:

$$APHR = (Number\ of\ Samples\ Correctly\ Classified)/(Total\ Number\ of\ Samples) \quad (1)$$

AHPR is the ratio of total correct classifications to the total number of samples, averaged for all classes in the classification problem and is more commonly known as precision. Precision is the percentage of films correctly assigned to their class, or in other words: "how many positive decisions by the classifiers are true?" The counterpoint to precision is fallout: "the proportion of non-class members that the system assigns to the class" (Lewis, 1997).

*2.6.11. Experimental results*

Prediction results were presented using two performance measures: average percent success rate of classifying a movie's success exactly (bingo), or within one class of its actual performance (one-away) (Delen & Sharda, 2010; Sharda & Delen, 2006). Initially, Sharda and Delen achieved a 36.9% bingo APHR with a purely ANN-based approach in 2006. Subsequently, they employed an expanded dataset and improved their ANN-based results to 52.5% and their fusion-based results to 56.07% (Table 5). The fusion model involves a voting scheme whereby a support vector machine model, a classification and regression tree model, a random forest model, a boosted tree model, and the original ANN model each provide a vote for the classification of each record and the class with the most votes is selected.

As previously mentioned, we use the Sharda and Delen model as a benchmark to show how well DAN2 performs against a leading, established approach by using identical data and performance metrics. We chose, however, to exclude the 'one-away' accuracy metric from our experiments because we do not believe this metric to be a sufficient measure of a model's accuracy. A 'bingo' classification is a more meaningful metric for decision making. For example, a correct 'one-away' classification for a record in class C means that a film has been classified as earning between $65 and $200 million. Because such a classification is not of substantive value to real world decision makers, we focus exclusively on the 'bingo' accuracy metric, excluding 'one-away' calculations from our replication of Sharda & Delen's experiments. We next present the data collection processes employed for the DAN2 models.

## 3. Data collection for DAN2 models

We were able to obtain a list of the titles of the films used by Sharda and Delen in their experiments, but the original datasets were not available in their entirety. We replicated 1758 of the 2632 movie records employed in their most recent experiment (Delen & Sharda, 2010). Although we were only able to reproduce 67% of this dataset, we used the same variables and methodologies employed by Sharda and Delen. Having fewer records on which to train an ANN for a classification problem can only increase the difficulty of the classification task. Sharda and Delen's ANN model initially achieved a classification accuracy of 36.9% with 834 records (Sharda & Delen, 2006). Subsequently, they more than tripled the size of their dataset by including of 1798 additional records, improved ANN model performance by 42.5% to an accuracy of 52.6% (Delen & Sharda, 2010).

Because some of the records utilized by Sharda and Delen were randomly missing due to data availability constraints, no manual pruning of the dataset occurred and no distinct advantage can be conferred by our having fewer records. Our data was procured from a variety of sources, including publicly available data, primarily from the-numbers.com and imdb.com. As previously stated, in addition to replicating the experiment performed by Sharda and Delen, we constructed an alternative model with more recent data to test the impact of production budgets, pre-release advertising expenditures, runtime, and seasonality on model performance for forecasting box-office revenue via the same classification method. We purchased our data from Kantar, a media research company.[1] This new dataset consists of 354 films which were released between 1999–2010.

In our second experiment, subsequent to our replication of the Sharda and Delen experiment, we employed the following variables: MPAA rating, sequel, screen count, production budget, pre-release advertising expenditures, runtime, and seasonality. Of these variables, production budget, pre-release advertising expenditures, and seasonality are not present in Delen and Sharda's experiment (2010). We evaluate the impact of these variables and demonstrate their positive contribution to model performance. These additional variables are described below.

### 3.1. Production budget

The production budget of a film has been consistently identified as a strong predictor of box-office performance (Basuroy & Chatterjee, 2008; Litman & Ahn, 1998; Pangarker & Smit, 2013). Although larger budgets are correlated with higher revenues, they

[1] Funding provided by Santa Clara University.

**Table 5**
Delen and Sharda (2010): prediction models & methods.

| | SVM | ANN | C&RT | Random forest | Boosted tree | FUSION (average) |
|---|---|---|---|---|---|---|
| APHR | 55.49% | 52.50% | 40.46% | 54.62% | 54.05% | 56.07% |

are not correlated with higher profits; and films with smaller budgets are, on average, more profitable (Basuroy et al., 2003). However, big budgets can act as an insurance policy in that it can serve to mitigate the negative effects of unfavorable critical reviews (Ravid & Basuroy, 2004).

### 3.2. Pre-release advertising expenditures

Advertising budgets, generally set in proportion to production costs, have jumped from roughly $10 million per film in the early 1990s to nearly $35 million per film in the early 2000s (MPAA, 2004). This increase may be attributable in part to television advertising price inflation coupled with the decreased reach of media networks (Hanssens et al., 2003). Advertising costs are the second largest expense for most films, with production budgets being the largest expense (Hanssens et al., 2003). The film industry has been measured as having among the highest advertising/sales ratio of all U.S. industries at 0.39 (Rennhoff & Wilbur, 2010). In addition to being strongly correlated to production costs, advertising expenditures are also so highly correlated with screen counts that the latter can be used as a proxy for the former (Hennig-Thurau et al., 2007).

Advertising strategy is tightly coupled with distribution strategy and advertising campaigns tend to peak when a film is released. On average, approximately 90% of advertising dollars are spent after production but prior to release (Elberse & Anand, 2007). An incremental (or 'platform') release schedule is sometimes employed as an alternative to a traditional advertising campaign in an attempt to generate word-of-mouth via limited exposure to the film itself rather than via widespread exposure to trailers and other media (Einav, 2007). Advertising campaigns generally begin 6–8 weeks before the release of a film (Hanssens et al., 2003). Because of this timing, the inclusion of advertising expenditure data in a predictive model limits the time during which a forecast can be conducted to the post-production period. We also offer a model without advertising data in order to provide decision makers a tool that can be used during the pre-production period, a time prior to the finalization of advertising budgets.

Advertising expenditures have been shown to be positively correlated with expected revenues in a statistically significant fashion via an analysis of movie stock prices on the Hollywood Stock Exchange prediction marketplace (Elberse & Anand, 2007). Despite this correlation, for the average movie, each dollar spent on advertising results in less than a dollar being earned as a result (Elberse & Anand, 2007).

### 3.3. Runtime

The runtime of a film (i.e. its duration in minutes), has been infrequently examined in the research, although no consistent conclusions have been drawn regarding its potential role as a predictive variable for the determination of box-office revenues. 'Art' films tend to have longer runtimes than blockbuster movies (Ainslie et al., 2005). Runtime is also more likely to influence the movie exhibitors' decision of whether or not to show a movie and less likely to influence individual movie goers' decision of whether or not to attend (Dhar et al., 2012).

### 3.4. Seasonality

Dramatic seasonal effects occur within the motion picture industry as the number of Americans who attend movies in theaters varies significantly over the course of the year, sometimes more than doubling within a two week period, generally around holiday weekends (Einav, 2007). The six holiday weekends most important to the U.S. motion picture industry are: President's Day, Memorial Day, The Fourth of July, Labor Day, Thanksgiving, and Christmas (Einav, 2007). During these times of peak demand, the films with the highest budgets are released, further amplifying underlying seasonality (Einav, 2007). Summer months, for instance, comprise both the most competitive release window and also the time period during which the highest box-office revenues are generally achieved (Fetto, 2010; Hennig-Thurau et al., 2007). Films released during seasons of peak attendance have theatrical runs of equivalent length to those released during the off-season (Radas & Shugan, 1998).

In addition to annual fluctuations in seasonal market effects, more than 75% of movies are released on Friday to compete for the weekend audience which brings in approximately 70% of all box-office revenues (Einav, 2007). Distributors' theatrical release timing continues to grow in importance as a strategic decision. Due to the increasingly interconnected nature of news and social media, the value of the network effects gleaned from being the most popular film at the box-office on an opening weekend continues to grow over time (Eliashberg et al., 2006). We include in our experiments a seasonality coefficient derived from approximately 2000 major film releases between 1985 and 2000 (Einav, 2007).

Hereafter we present a short description of DAN2 and offer an alternative model for movie forecasting using the variable set which has just been introduced.

## 4. DAN2 model

We have developed a Dynamic Architecture for Artificial Neural Networks (DAN2), which employs a different architecture than the traditional neural network (FFBP) models. The general philosophy of the DAN2 model is based upon the principle of learning and accumulating knowledge at each layer, propagating and adjusting this knowledge forward to the next layer, and repeating these steps until the desired network performance criteria are reached (Fig. 1).

As in classical neural networks, the DAN2 architecture is composed of an input layer, hidden layers and an output layer. The input layer accepts external data to the model. The $(n \times m)$ matrix in Fig. 1 presents the input layer. This layer constitutes the $n$ training nodes with each node representing an observation vector with $m$ attributes. For space brevity, a matrix representation is used in Fig. 1 rather than the $n$-nodes. Furthermore, the node $I$ is used to present this input set, thus making the matrix and the node $I$ to be same. In DAN2, unlike classical neural nets, the number of hidden layers is not fixed a priori. They are sequentially and dynamically generated until a level of performance accuracy is reached. Additionally, the proposed approach uses a fixed number of hidden nodes (four) in each hidden layer. This structure is not arbitrary, but justified by the estimation approach. At each hidden layer, the network is trained using all observations in the training set simultaneously, so as to minimize a stated training accuracy measure such as mean squared error (MSE) value or other accuracy measures. As shown in Fig. 1, each hidden layer is composed of four nodes. The first node is the bias or constant (e.g. 1) input node, referred to as the *C* node. The second node is a function that encapsulates the "Current Accumulated Knowledge Element" (CAKE node) during the previous training step. The third and fourth nodes represent the current residual (remaining) nonlinear component of the process via a transfer function of a weighted and normalized sum of the input variables. Such nodes represent the "Current Residual Nonlinear Element" (CURNOLE nodes). In Fig. 1, the "*I*" node represents the input, the "*C*" nodes are the constant nodes, the "$G_k$" and "$H_k$" nodes represent CURNOLE nodes, and the "$F_k$" nodes are the CAKE nodes. The final CAKE node represents the dependent variable or the output.

At each layer, the previous four nodes ($C$, $G_k$, $H_k$, and $F_{k-1}$) are used as the input to produce the next output value ($F_k$). The parameters on the arcs leading to the output nodes ($a_k$, $b_k$, $c_k$, $d_k$) represent the weights of each input in the computation of the output for the next layer. The parameter connecting the CURNOLE nodes, $\mu_k$, is used as part of the argument for the CURNOLE nodes and reflects the relative contribution of each input vector to the final output values at each layer. A detailed description of the architecture and its properties are fully presented in (Ghiassi & Saidane, 2005).

The training process begins with a special layer where the CAKE node captures the linear component of the input data. Thus, its input (content) is a linear combination (weighted sum) of the input variables and a constant input node. These weights are easily obtainable through classical linear regression. The $F_0$ in Fig. 1 presents the starting point and is often computed using the input matrix and ordinary linear regression. If the desired level of accuracy is reached, we can conclude that the relationship is linear and the training process stops. This step is used as the starting point. For classification problems this step can be replaced with alternative methods described in Section 4.12. For nonlinear relations additional hidden layers are required. At each subsequent layer the input to the CAKE node is a weighted sum (linear combination) of the previous layer's CAKE, CURNOLE, and C nodes. Therefore, after the first iteration, (the starting point), the input to DAN2 includes the $F_0$ value, a constant term and an arbitrary value for $\mu_1$ (e.g. $\mu_1 = 1$). These parameter values are used to compute $F_1$ as described below (Eqs. (2) and (3)). Throughout training, the CAKE nodes carry forward an adequate portion of learning achieved in previous layers. This process ensures that the performance or knowledge gained so far is adjusted and improved but not lost. This property of DAN2 introduces knowledge memorization to the model. Ghiassi and Saidane (2005) show that the DAN2 algorithm ensures that during network training, the residual error is reduced in every iteration and the accumulated knowledge is monotonically increased.

The training process defines creation of partitions among classes that could include linear and nonlinear components. The linear component of the input data is captured in the first CAKE node using ordinary least squares (OLS) or other simple and easy to compute approaches. The algorithm next transforms the input dataset to model the nonlinearity of the process in subsequent iterations. DAN2 uses a vector projection approach to perform data transformation. The transformation process defines a reference vector $R = \{r_j;\ j = 1, 2, \dots m\}$, where $m$ represents the number of attributes of the observation records, and projects each observation record onto this vector to normalize the data as discussed in (Ghiassi & Saidane, 2005). This normalization defines an angle, $\alpha_i$, between record $i$ and the reference vector R. DAN2 uses the set of $\alpha_i$'s to train the network, and updates their values in every iteration. At each layer, the training (observation) vectors are projected onto the reference vector R. Their collective influence contributes to the knowledge gained at that iteration. In the subsequent iterations, the reference vector R is "intelligently" moved (rotated and/or shifted). This movement changes the angle $\alpha_i$ and thus revises the contribution of each vector to the cumulative knowledge. The challenge of the algorithm is directly related to how to "intelligently" move the reference vector for each iteration. In Ghiassi and Saidane (2005), the authors show that this process can be represented by the trigonometric function Cosine ($\mu_k\alpha_i + \theta_k$) – a normalization process. In every hidden layer $k$ of the architecture the reference vector is moved thus changing $\alpha_i$, as represented by ($\mu_k\alpha_i + \theta_k$). This move measures the impact of this change on the output value. The modification of the angle ($\mu_k\alpha_i + \theta_k$) is equivalent to rotating $\mu_k$ and shifting $\theta_k$, the reference vector, thus changing the impact of the projected input vectors and their contribution to the output for that iteration. This transformation uses the Cosine ($\mu_k\alpha_i + \theta_k$) function with two (nonlinear) parameters, $\mu_k$ and $\theta_k$. To simplify the resulting optimization process, the use of the latter can be avoided through the expansion of the cosine function in the general form: *A* Cosine ($\mu_k\alpha_i$) + *B* Sine ($\mu_k\alpha_i$). We use this functional form as the transfer function in our model. The two CURNOLE nodes in Fig. 1 represent this formulation. At any given hidden layer $k$, if the Cosine ($\mu_k\alpha_i + \theta_k$) terms captured in previous layers do not adequately express the nonlinear behavior of the process, a new layer with an additional set of nodes is automatically generated, including a new Cosine ($\mu_k\alpha_i + \theta_k$) term. This process is analogous to how the Fourier series adds new terms to improve function approxi-

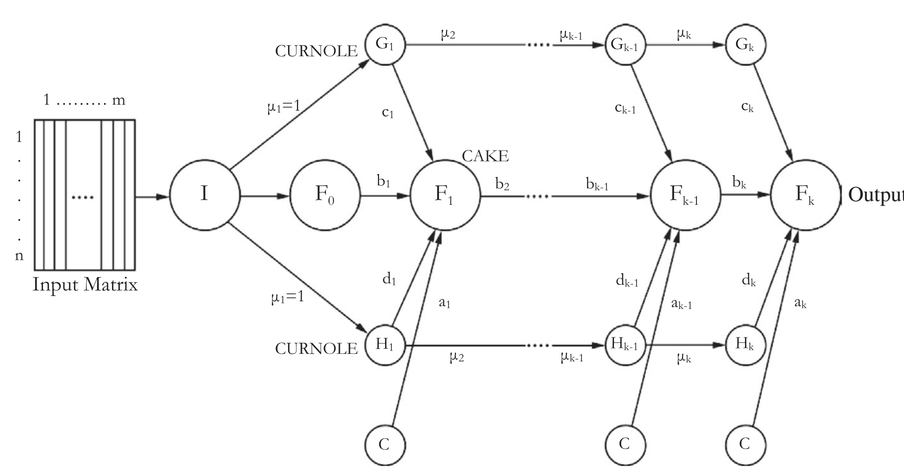


**Fig. 1.** DAN2 network architecture.

mation. Therefore, the number of layers in the DAN2 architecture is dynamically defined and depends upon the complexity of the underlying process and the desired level of accuracy. Thus, the output of this model is represented by the linear combination of the constant, CAKE, and CURNOLE nodes. Eq. (2) represents the functional form of this relationship at iteration (layer) $k$

$$F_k(X_i) = a_k + b_k F_{k-1}(X_i) + c_k G_k(X_i) + d_k H_k(X_i) \tag{2}$$

where $X_i$ represents the $n$ independent input records, $F_k(X_i)$ represents the output value at layer $k$, $G_k(X_i)$ = Cosine ($\mu_k \alpha_i$), and $H_k(X_i)$ = Sine ($\mu_k \alpha_i$) represent the transferred nonlinear components, and $a_k$, $b_k$, $c_k$, $d_k$, and $\mu_k$ are parameter values at iteration $k$. The training process initially captures the linear component by using OLS or other simple and easy to compute approaches. If the desired level of accuracy is reached, the training terminates. Otherwise, the model generates additional layers to capture the nonlinear component of the process by minimizing a measure of total error as represented by $SSE_k = \Sigma_i[F_k(X_i) - F^{\wedge}(X_i)]^2$. Substituting $F_k(X_i)$ from Eq. (2) results in:

$$SSE_k = \sum_i \left[a_k + b_k F_{k-1}(X_i) + c_k \cos(\mu_k \alpha_i) + d_k \sin(\mu_k \alpha_i) - F^{\wedge}(X_i)\right]^2 \tag{3}$$

where $F^{\wedge}(X_i)$ are the observed output values. Minimizing Eq. (3) requires the estimation of five parameters. This formulation is linear in the parameter set $A_k$, where $A_k = \{a_k, b_k, c_k, d_k\}$ and nonlinear in parameter $\mu_k$. In Ghiassi and Saidane (2005), they present several nonlinear optimization strategies to estimate the nonlinear parameter $\mu_k$. They also show that following this approach, at each layer the knowledge gained is monotonically increased, total error is reduced, and the network training improves. The authors in Ghiassi and Saidane (2005) present several approaches to solve the optimization problem stated by Eq. (3). The approaches recommended include a gradient-based optimization approach as well as a grid-search based approach. In each approach the solution requires estimation of the linear parameter $A_k = \{a_k, b_k, c_k, d_k\}$ and the nonlinear parameter $\mu_k$. They offer a layer-centered training process, and decompose the solution to linear and nonlinear stages to increase the efficiency of solving (minimizing) the highly nonlinear Eq. (3). The entire training dataset is collectively used in both the linear and nonlinear stages. In the linear stage, a simple OLS is used to estimate the parameter set $A_k$ for a given value of $\mu_k$. The second stage uses the estimated values of $A_k$ and searches in the nonlinear space of $\mu_k$ for the value of $\mu_k{}^*$ that minimizes $\text{SSE}_\text{k}$. The optimal value of $\mu_k$ alters the product ($\mu_k{}^* \alpha_i$) and changes the contribution of input vector $X_i$ to the output through the Cos($\mu_k{}^* \alpha_i$) and Sin($\mu_k{}^* \alpha_i$) components. Once the best value of $\mu_k$ for layer $k$ is determined, stopping rules are examined. If the stopping criteria are met, training terminates; otherwise, a new layer is introduced and the process repeats.

The approach used in this implementation is based on the grid-search approach. We note that the grid-search approach is a heuristic method that does not guarantee optimality. Therefore, the algorithm may discard evaluation layers that do not improve the objective function value and will try another value instead. We also note that the value of $K$ represents all the trials layers in this approach, including the discarded layers, and thus the value of $K$ lacks significance. In comparison, the gradient-based search approach only generates layers that will improve objective function values and thus the value of $K$ would represent the total number of layers in the network.

Similar to other neural network algorithms, DAN2 is iterative in nature and reaching the final iteration is always experimental and relies on the user defined accuracy metrics. The two metrics often used are: (1) a user defined measure of accuracy, and (2) the total number of iterations. Additionally, such algorithms also offer some internal measures to guard against under/over fitting concerns. DAN2 follows the same principles for both model fitting and the avoidance of under/over fitting. In the grid-search approach used in this implementation, a lower accuracy value is first used to identify promising search intervals. Once good candidate intervals are identified, the accuracy level and the maximum number of iterations are increased.

To avoid over-fitting during the model selection process we use a simplified "cross-validation" approach. We use a representative fraction of the in-sample data (10–20%) for validation and to compute the over-fitting metric $\varepsilon_2$ (defined below) after every iteration. Various strategies can be used to select the validation set. The strategies may range from selecting the last 10–20% of the in-sample dataset, or for cyclic processes, at least one full cycle of the data, or a randomly selected set of records from the in-sample dataset. If model training stops prematurely, the network is considered to be "under-trained" or "under-fitted." An under-trained model often has high SSE values for either or both the training and validation datasets. Under-training often occurs when there are insufficient data for model fitting. We use $\varepsilon_1 = (\text{SSE}_{k-1} - \text{SSE}_k)/\text{SSE}_{k-1} \leqslant \varepsilon_1^*$ to assess existence of under-training in our models at iteration $k$. Over-training or over-fitting is a more common problem in neural net modeling. A neural net model is considered over-fitted (over-trained) when the network fits the in-sample data well but produces poor out-of-sample forecasts. To avoid over-fitting, at each iteration $k$, ($k > 1$), we compute MSE values for both the training ($\text{MSE}_\text{T}$) and validation ($\text{MSE}_\text{v}$) datasets. We use $\varepsilon_2 = |\text{MSE}_\text{T} - \text{MSE}_\text{v}|/\text{MSE}_\text{T} \leqslant \varepsilon_2^*$ to guard against over-fitting. The model is considered fully trained when the user specified accuracy criteria and the over-fitting constraint are both satisfied. When a user specifies a relatively small value for the accuracy measure, it may be possible to reach the over-fitting criterion $\varepsilon_2$ before reaching the desired level of accuracy. The modeler then needs to reexamine and revise the desired level of accuracy in order to avoid over-fitting.

DAN2's algorithm is highly scalable. The algorithm is composed of a series of layers that are automatically and dynamically generated. At each layer the observation vectors are projected into the reference vector to create the angle $\alpha_i$ as discussed earlier. The training and optimization at each layer only uses the transformed data to estimate the five parameters of Eq. (3). Therefore, regardless of the original size of the dataset, DAN2 only needs to compute the set of four linear parameters, $A_k = \{a_k, b_k, c_k, d_k\}$, and one nonlinear parameter $\mu_k$ at each iteration. For instance, for text classification problems with large feature sets, once the starting point is computed, each observation record, as represented by its many features, is projected onto the reference vector and its angle, $\alpha_i$, is computed. The training process only uses the $\alpha_i$ values and only needs to compute the five parameters. Therefore, the original problem with a large feature set is converted to a 5-parameter model at each layer; thus scaling the problem size and demonstrating the scalability of the model. The scalability of DAN2 is a distinguishing strength of the approach from traditional artificial neural networks.

In previous research (Ghiassi & Burnley, 2010; Ghiassi, Olschimke, Moon, & Arnaudo, 2012; Ghiassi, Saidane, & Zimbra, 2005) the authors compare DAN2 with traditional FFBP and recurrent neural network (RNN) models. The comparison spans both theoretical and computational perspectives using several benchmark datasets from the literature. Performance of DAN2 against these models as well as non-neural network alternatives is also presented. Their studies show that DAN2 outperforms all other alternatives by producing more accurate training and testing results in every case.

### 4.1. Using DAN2 for multiple class classification

The movie forecasting model in this research transforms the problem into a multiple class classification. The additional complexity inherent in multiple class classification presents a challenge to many classification algorithms. However, in order to reduce this complexity, almost all the algorithms break the problem into a one-versus-all classification. In this approach for a multiple class problem of size n, n models are developed with each model representing a class. DAN2 also uses the one-versus-all approach to classify movies into their corresponding nine classes. Similar to other algorithms, the results are reconciled by classifying a sample as belonging to the version in which it has the highest output value. The $F_k$ value in DAN2 provides this output value for each model. We note that the partitioning suggested in this implementation relies on availability of adequate data for each class. When developing the model for each class, the training and testing datasets consists of the entire available data. When a class has very few members, and it is disproportionally small, referred to as an unbalanced class, care must be taken to ensure proper training of the model. For classification models, accuracy metrics are often simple measures of whether or not a record belongs to a given class. Accordingly, when the favorable instances (the records belonging to a class) are very few (e.g. less than 5% of the total dataset), if the algorithm simply assigns all the records as "not belonging" to the class, the results would technically be 95% accurate. This is the well-known "unbalanced" problem in classification and to avoid such scenarios modelers may randomly reduce the unfavorable records of the training dataset in order to increase the ratio of favorable records to an acceptable level. In this research the number of favorable records for all models using DAN2 was sufficient and thus no balancing of the datasets was required. Finally, the alternative to the one-versus-all solution for multiple class classification is the hierarchical approach. DAN2 offers such an option and the details of its hierarchical approach are described in Ghiassi and Burnley (2010).

### 4.2. DAN2 operation

Creating a good starting point in any nonlinear optimization problem is always challenging. However, good starting point values often improve nonlinear optimization convergence. For this analysis, we employ a basic K-Nearest-Neighbors (KNN) algorithm to provide an easily calculable starting point for DAN2. During this process we calculate results for $k$ values from 1 to 6, select the best results as our starting point, and proceed with DAN2 classification.

All machine learning algorithms, including DAN2, require experimentation to find the best trained model. DAN2 initially defines a set of stopping conditions, including the maximum number of iterations and the target accuracy (the algorithm will stop at whichever condition is met first). At every iteration DAN2 solves a nonlinear optimization problem. In Ghiassi and Saidane (2005), the authors offer a number of alternatives to solve the intermediate nonlinear optimization problem. The implementation used in this research utilizes the grid-search approach. In DAN2 a 2-dimensional search (interval and step sizes) is employed to find the best parameter value at each iteration. The algorithm begins with larger intervals and step sizes to determine promising search regions. When promising sectors are identified, subsequent more granular 2-D searches are performed on each sector to find the (local) optima. When a step produces inferior results, it is discarded and the process starts over with a different interval and step size. Once the desired level of training accuracy is reached, the search terminates. The algorithm uses internally set metrics to avoid over-fitting as described in Ghiassi et al. (2005).

Since there are sufficient records for both the model employed by Sharda and Delen and the DAN2 model to be thoroughly trained, we conclude that any performance difference can solely be attributed to the tool used to capture the behavior of the data.

### 4.3. Accuracy metrics

In addition to the APHR employed by Delen and Sharda (2010), we include $F_1$ scores as a more robust measure of ANN model classification accuracy. An $F_1$ score is a calculation of the trade-off between two component metrics: precision and recall (Eq. (4)).

$$F_1 = 2 * (precision * recall)/(precision + recall) \tag{4}$$

Precision, as previously discussed, is the ratio of total correct classifications to the total number of samples. Recall, on the other hand, is a measure that provides this information: it "is the proportion of class members that the system assigns to the class," (Lewis, 1997). In other words, of all positive expert judgments, how many have been "found" by the classifier? A perfect classifier would have an accuracy, precision and recall of 1.0 and error-rate and fallout of 0.0 (Lewis, 1997). However, both quantities trade off against each other: it is easy to achieve a recall of 1.0 but very low precision, just by classifying all films into a given class (Manning, Raghavan, & Schütze, 2008). The precision/recall break-even point is the point where both metrics are in balance (Manning et al., 2008). This metric is widely used as a metric for classification accuracy (Joachims, 1999). A measure of the trade-off between precision versus recall is the $F$ measure, commonly written as the $F_1$ measure (Manning et al., 2008).

## 5. Box-office forecasting

DAN2 has been shown to be a highly effective tool for classification (Ghiassi & Burnley, 2010; Ghiassi et al., 2012) and forecasting problems (Ghiassi & Nangoy, 2009; Ghiassi et al., 2005). We first use the exact same variable set and accuracy metrics employed by the two Sharda and Delen models with DAN2. Side by side comparison of DAN2 with Sharda and Delen's models allow us to demonstrate the effectiveness of DAN2 as a modeling approach for box-office revenue forecasting and classifications. We then present our approach to modeling pre-production box-office revenue forecasting using DAN2 and show that the alternative variable set presented and the utilization of DAN2 allows us to reach a high degree of accuracy, thus providing stakeholders with a decision support system for analyzing investment opportunities in this volatile domain.

### 5.1. DAN2 versus Sharda and Delen ANN

Sharda and Delen's experiments employ 26 variables and 2632 movies using accuracy metrics introduced earlier. We use the exact same 26 variables but, as indicated in Section 3, we could only obtain data for 1758 movies. We train our model using 80% of these records and then test it using the remaining 20%. In addition to using the same metrics employed by Sharda and Delen, we also use a more widely accepted accuracy metric for classification problems, the $F_1$ score (Manning et al., 2008). The close performance values between the training and testing datasets is evidence that our model is properly trained and is neither over nor under fitted. We also can conclude that the smaller dataset used by DAN2 was sufficient for training and testing and the availability of more data (randomly distributed throughout the dataset) could have only contributed positively to the performance of DAN2.

By employing DAN2 to the same data with the same methodology, we were able to improve classification accuracy from the

**Table 6**
DAN2 results using Sharda and Delen benchmark experiment.

| Class | Training | | Testing | |
|---|---|---|---|---|
| | APHR (%) | F1 (%) | APHR (%) | F1 (%) |
| A | 94.60 | 72.60 | 66.70 | 65.30 |
| B | 97.20 | 80.10 | 75.00 | 78.10 |
| C | 58.50 | 66.70 | 61.90 | 71.00 |
| D | 35.20 | 72.30 | 58.10 | 67.30 |
| E | 47.90 | 60.90 | 54.80 | 67.60 |
| F | 65.30 | 46.70 | 61.70 | 69.20 |
| G | 60.40 | 69.10 | 66.00 | 74.30 |
| H | 85.20 | 92.10 | 90.10 | 80.00 |
| I | 64.30 | 93.30 | 62.00 | 80.00 |
| Average | 73.50 | 70.60 | 74.40 | 71.30 |

56.01% APHR benchmark previously established by Sharda and Delen to 74.4% APHR, an improvement of 18.39 percentage points, or 32.83% (Table 6). Additionally, $F_1$ scores of 70.6% training and 71.3% testing were achieved and are presented here as more robust benchmarks for future experiments. Individual class metrics are not available for Sharda and Delen's results, nor is information regarding training versus testing accuracy available. We attribute the classification performance improvement to the DAN2 algorithm. This analysis demonstrates that DAN2 is a more effective tool for box-office revenue forecasting. We acknowledge that this conclusion is based upon one box-office revenue dataset. We conclude that these results provide evidence that DAN2 performs better without claiming either generality or superiority. Generalization of such a claim requires the application of DAN2 to a larger set of identical box-office problems.

### 5.2. DAN2 pre-production forecasting model

We next introduce a different model using a different set of variables and a dataset representing a more recent set of movies. We continue to use $F_1$ as an accuracy metric. The main focus of this model is to introduce a set of variables with values that are available early in the life-cycle of movie production. This allows for the development of a decision support tool that can assist decision makers early in the movie production process. The DAN2 pre-production forecasting model diverges from the previous model by including a different set of variables (Table 7). Competition, star value, genre, and special effects have been removed as variables, while production budget, pre-release advertising expenditures, runtime, and seasonality have been added.

Competition was included in Sharda and Delen's experiments as a measure on a scale from 1 to 3 of the variety, depth, and prominence of contemporaneously alternative films. Instead, we include as a proxy variable, an independently developed seasonality coefficient, of which competition is a component factor (Einav, 2007).

Star power is almost always highly correlated with production budget (Basuroy et al., 2003; Hennig-Thurau et al., 2007). As discussed previously, prior research is also highly divided as to the predictive potential of star power for box-office revenue forecasting. Furthermore, no uniform method of objectively calculating star power exists. Accordingly, we believe that the inclusion of production budgets (in dollars) can also be used as a proxy variable for star power.

Because movies can belong to more than one genre and no uniform set of genre classifications are broadly benchmarked in prior research, exclusion of genre variables reduces model complexity. We include runtime in minutes as a pseudo-proxy variable for genre, given that 'art' films tend to have longer runtimes than blockbuster movies (Ainslie et al., 2005).

The special effects variable is excluded due to the difficulty involved with calculating a consistent, objective measure for the level of special effects in a film. No such methods have been published in box-office forecasting literature. Additionally, no prior research exists that would suggest special effects levels are directly correlated with film revenues, nor is requisite data publicly available.

The inclusion of pre-release advertising expenditures in the DAN2 pre-production forecasting model is based upon a previously demonstrated positive correlation. Advertising expenditures and expected revenues are related in a statistically significant fashion, as observed in prior analyses of movie stock prices on the Hollywood Stock Exchange prediction marketplace (Elberse & Anand, 2007).

By removing competition, star value, and special effects from our experiment, we are able to extend backwards the available timeline during which a revenue forecast with DAN2 is viable. Advertising budgets, being largely spent during the post-production, pre-release time frame, are also highly subject to fluctuation from initial estimates made during the pre-production period. Additionally, having observed no need to include advertising expenditure data, we avoid the expense and effort needed to acquire such data, which is no longer publicly available nor easily calculable (due to fractured media channels). By including only variables which can be known or assigned prior to the production of a film, we position DAN2 as a viable pre-production decision support tool for revenue forecasting.

We were able to obtain pre-release advertising data for 354 films. Following the general practices employed in ANN modeling, we divided that data: 283 (80%) for training and 71 (20%) for testing. Additionally, we have implemented new class thresholds (Table 8), narrowing revenue ranges and achieving more meaningfully accurate classifications.

Our classification methodology remained the same for this dataset. We performed two experiments, one including the pre-release advertising budget variable and one excluding this variable. While both experiments achieve classification rates higher than those achieved with the Sharda and Delen dataset, we observed no significant improvement when advertising data was included. Confirming the findings of earlier researchers, we attribute this to advertising expenditures being highly correlated with both

**Table 7**
DAN2 model inputs.

| Variable | Possible values |
|---|---|
| MPAA rating | G, PG, PG-13, R, NR |
| Sequel | True, false |
| Number of screens | Positive integer |
| Production budget (dollars) | Positive integer |
| Pre-release advertising expenditures (dollars) | Positive integer |
| Runtime (minutes) | Positive integer |
| Seasonality | 0.65–1.20 |

**Table 8**
DAN2 model class thresholds.

| Class | Range (in $ millions) |
|---|---|
| (Blockbuster) A | 100–200 |
| B | 80–100 |
| C | 60–80 |
| D | 40–60 |
| E | 30–40 |
| F | 20–30 |
| G | 10–20 |
| H | 1–10 |
| (flop) I | <1 |

**Table 9**
DAN2 Pre-production forecast results.

| Class | Count | DAN2 (without advertising) | | DAN2 (with advertising) | | F1 testing difference (with ad. – without ad.) (%) |
|---|---|---|---|---|---|---|
| | | F1 training (%) | F1 testing (%) | F1 training (%) | F1 testing (%) | |
| A | 21 | 100.00 | 85.70 | 100.00 | 88.90 | 3.20 |
| B | 21 | 97.30 | 100.00 | 100.00 | 100.00 | 0.00 |
| C | 28 | 100.00 | 88.90 | 100.00 | 88.90 | 0.00 |
| D | 37 | 100.00 | 92.30 | 100.00 | 93.30 | 1.00 |
| E | 44 | 97.80 | 100.00 | 100.00 | 100.00 | 0.00 |
| F | 32 | 94.30 | 100.00 | 96.30 | 100.00 | 0.00 |
| G | 46 | 86.50 | 93.30 | 78.90 | 85.70 | −7.60 |
| H | 62 | 81.30 | 85.70 | 83.90 | 87.00 | 1.20 |
| I | 63 | 100.00 | 100.00 | 100.00 | 94.70 | −5.30 |
| Total/Avg. | 354 | 94.00 | 94.10 | 94.10 | 92.70 | −1.40 |

production budgets and screen counts (Hennig-Thurau et al., 2007). In other words, since we are already including the production budget and screen count at release for a given film, we are not adding significant predictive information by including the highly correlated variable of pre-release advertising expenditures.

Table 9 presents the results of box-office revenue forecasting with and without advertising values for the nine movie classes. The overall training and forecasting accuracy values ($F_1$ values) for the "without advertising" models, weighted by each class, are 94.0% & 94.1% respectively. The best results reach 100% accuracy for the B, E, F, & I models (using the testing accuracy measure). The most volatile classes are the A and H classes with testing accuracy level of 85.7%.

Class A, the 'blockbuster' class, contains films with the largest budgets and the highest revenues. This class may be slightly underperforming other classes due to the difficulty associated with forecasting highly successful yet low-budget films. Additional contributing factors may include the scope of the revenue for class A. Future experiments, not limited by scarce advertising expenditure data, would provide further information and allow for narrower class ranges in the 'blockbuster' range. Alternatively, Class H (revenue of between 1 and 10 million USD), may be underperforming due to the internal spectrum breadth of this class' revenue range. No other class has the potential to contain films separated in revenue by an order of magnitude (i.e. 1 million USD and 10 million USD). The nearest another class comes to comparability on this metric is class A, in which only a doubling is possible (i.e. 100 million USD and 200 million USD).

Forecasting results for a majority of classes (7 out of 9) are at or above 90%. These results are excellent and compare very favorably against any published reports. The uniformly excellent performance of this model validates the choice of the new variable sets for box-office revenue forecasting models along with the effectiveness of DAN2 as a classification tool. The close accuracy values for the training and testing datasets reflect the absence of over/under fitting.

Table 9 also presents the results of box-office revenue forecasting when we add the pre-release advertising expenditure variable to the model. While training metrics for some classes improved slightly, all classes either decreased in testing accuracy or remained static.

Although advertising expenditures have been shown to be positively correlated with expected revenues, our research confirms the earlier findings of (Elberse & Anand, 2007) that pre-release advertising expenditures do not offer additional predictive value. We observed a decrease in overall, weighted average testing $F_1$ of 1.4% (94.1–92.7%). This result is attributable at least in part to advertising budgets' general proportionality to production costs (MPAA, 2004). Advertising expenditures are also so highly correlated with screen counts that the latter can be used as a proxy for the former (Hennig-Thurau et al., 2007). Because our model includes both production budgets and screen counts, we conclude that insufficient additional information content is contained within pre-release advertising expenditures to positively influence forecasting accuracy.

Neural network models are not generally used to establish causation because there are no coefficients unique to particular variables. However, neural network models can use the pruning or additive modeling approaches to assess the contribution of a variable to forecasting accuracy. Subsequent to our initial replication of Sharda and Delen's experiments, we present two additional models, one that includes pre-release advertising expenditure figures and one that does not include this variable. By observing the performance difference resulting from the inclusion of pre-release advertising expenditures, we conclude that for our dataset this is not a variable of significance to the task of box-office revenue forecasting.

To provide a point of comparison, we ran the model without advertising expenditures with Support Vector Machines (SVM). Table 10 lists those results. When running SVM, our first batch used the exact same training and testing datasets that were used for the DAN2 runs (using roughly an 80/20 split between out-of-class and in-class records). These datasets were characterized by a scarcity of in-class records to reflect the overall representation in the whole dataset. In a number of cases, though, this was modified to provide 20% of records in-class and 80% out-of-class. As indicated in the table, SVM struggled greatly with this distribution, and in all but two classes simply classified everything in the test dataset as out-of-class. Because SVM is designed for binary classification problems, we used the same one-vs-all approach with SVM as was used with DAN2. We used the Weka platform to perform SVM experiments, employing the standard polynomial kernel with default settings. The SVM package employed in this paper uses the Polynomial Kernel. This kernel was used with an exponent value of 1 and a cache size of 250,007 (both of these are default values). The complexity constant used is 1.0 (also the default value).

Given this poor performance, we modified the training datasets for each class to provide something more balanced. When we made the balance 33% in-class and 67% out-of-class, there was some improvement, but not enough to make this a viable model. The final test we did had a perfect balance between in-class and out of-class-records. When we did this, SVM did a much better job of classifying. The overall results, though, are still significantly less than DAN2. A significant contributing factor to this difference may be the small size of the dataset. Future experiments with a larger dataset may provide better information to explain this difference in performance.

With the DAN2 pre-production forecasting model accuracy on a per class basis ranged from 86–100%. We achieved a weighted average classification rate of 94.1%, with individual class accuracies

**Table 10**
DAN2 and SVM pre-production forecast results (without advertising expenditures).

| Class | Count | DAN2 F1 test (%) | SVM F1 test | | |
|---|---|---|---|---|---|
| | | | 80/20 Split (%) | 67/33 Split (%) | 50/50 Split (%) |
| A | 21 | 85.70 | 0.00 | 0.00 | 30.80 |
| B | 21 | 100.00 | 33.30 | 40.00 | 53.30 |
| C | 28 | 88.90 | 0.00 | 0.00 | 28.60 |
| D | 37 | 92.30 | 0.00 | 0.00 | 43.80 |
| E | 44 | 100.00 | 0.00 | 28.60 | 22.20 |
| F | 32 | 100.00 | 0.00 | 28.60 | 41.70 |
| G | 46 | 93.30 | 0.00 | 0.00 | 40.00 |
| H | 62 | 85.70 | 0.00 | 26.10 | 26.10 |
| I | 63 | 100.00 | 75.00 | 80.00 | 80.00 |
| Average | 354 | 94.10 | 15.30 | 27.30 | 42.40 |

weighted in proportion to their constituency in the total dataset (Table 9). The improvement in model accuracy is attributable in part to our revised variable selection and model construction. These results further demonstrate DAN2 to be an effective tool for box-office revenue forecasting. Additionally, our modeling methodology allows for decision support at the genesis of the movie production lifecycle, well in advance of competing models.

### *5.3. DAN2 computational times*

All machine learning algorithms, including DAN2, ANN, kNN, and SVM, require experimentation to find the best training model. For example, for SVM, the value of the penalty parameter and the choice of the kernel function needs to be selected experimentally (Joachims, 1998), while with kNN, the best $k$ value can only be determined by trial and error. For ANN models, the architecture (number of layers and number of hidden nodes) is experimentally defined, and then the parameters of the model are computed. Similarly, training of a DAN2 model requires experimentation. DAN2 initially defines a set of stopping conditions. These conditions include the maximum number of iterations and the target accuracy (the algorithm will stop at whichever condition is met first). At each iteration, DAN2 solves a nonlinear optimization problem. The implementation used in this research utilizes the grid-search approach. In DAN2 a 2-dimensional search (interval and step sizes) is employed to find the best $\mu_k$ at each iteration. The algorithm begins with larger intervals and step sizes to determine promising search regions. When promising sectors are identified, subsequent more granular 2-D searches are performed on each interval to find the (local) optima. When a step produces inferior results, it is discarded and the process starts over with a different interval and step size. Once the desired level of training accuracy is reached, the search terminates. The algorithm uses internally set metrics to avoid over-fitting as described earlier.

We timed the computational effort required to train each class to assess DAN2's efficacy. Table 11 reports training and testing times required for each of the nine models. We note that although the size of the training set is an important variable in model training time, it is not the only significant factor. Clearly, the complexity of the problem, the underlying nonlinear model and the starting points, are also influential. As presented in Table 11, in our computational environment, the two classes at the ends of the range (blockbusters and flops) took the longest to train (16.257 s and 39.784 s respectively). Once the models are trained, however, the actual runtime of the model for the testing sets is very short. For this dataset, the longest testing time for the trained DAN2 models is 0.122 s for class I (flops, <$1 M) and the shortest time is 0.016 s for class H ($1 M to $10 M). Similar information for SVM, and Sharda & Delen's ANN was not provided in the literature and thus, we cannot directly compare DAN2's computational times with others.

**Table 11**
DAN2 training and testing times.

| Class | DAN2 training time (s) | DAN2 testing time (s) |
|---|---|---|
| A | 16.257 | 0.055 |
| B | 3.126 | 0.053 |
| C | 8.502 | 0.040 |
| D | 7.045 | 0.042 |
| E | 7.673 | 0.046 |
| F | 3.593 | 0.018 |
| G | 2.477 | 0.038 |
| H | 2.861 | 0.016 |
| I | 39.784 | 0.122 |

## 6. Conclusions

Significant capital and time investments go into the production of every major motion picture. However, the majority of these films earn less from gross domestic box-office revenues than their production costs. Previously established models have focused primarily upon post-release and post-production forecasts. These models often rely upon opening weekend data and are reasonably accurate, but only if data up until the moment of release is included (Foutz & Jank, 2007). A forecast made immediately prior to the debut of a film, however, is of limited use to stakeholders because it can only influence late-stage adjustments to advertising or distribution strategies and little else.

In this paper we have presented the development of a model based on DAN2 for forecasting movie revenues during the pre-production period. We first demonstrated DAN2 to be an effective modeling tool for movie revenue forecasting. We used DAN2 to replicate published experiments and improved upon benchmark model performance by 32.8%. This contribution should encourage researchers to employ DAN2 as an effective ANN tool for movie revenue forecasting.

Subsequently, we offered an alternative modeling strategy that is shown to offer a superior forecasting accuracy of 94.1%. This modeling strategy utilized a new variable set. In the alternative model, competition, star value, genre, and special effects were removed as variables, while production budget, pre-release advertising expenditures, runtime, and seasonality were added. Although an ANN is not an effective tool for establishing or measuring causation, our experimentation has resulted in a variable set that has produced remarkable accuracy at 94.1%. We therefore conclude that the choice of the variables combined with the efficacy of DAN2 validates our model.

The box-office revenue forecasting model introduced in this research and its superior accuracy establish an excellent decision support tool for stakeholders in the movie industry, offering them a novel, practical advantage. This is especially true during the pre-production period when requisite data for alternative models are generally unavailable.

Advertising budgets are the second largest expense for most films, with production budgets being the largest. Additionally, advertising data is protected industry information and its sources are often proprietary. Another contribution of this paper is the collection (purchase) and inclusion of advertising data in our model. Researchers have previously inquired as to the role of this data in box-office revenue forecasting. We have developed separate models using both DAN2 and SVM to measure the impact of pre-release advertising expenditures on box-office revenue. The inclusion of this variable did not result in significant improvement in revenue forecasting. This lack of effect is attributed to the well-established positive correlation between advertising expenditures with both production costs and screen counts, two variables also included in our model. Due to the difficulty involved with acquiring this data and its proprietary nature, these results should be of value to future researchers.

The approach followed in this paper classified movies into nine groups with each group representing a revenue interval. Future research will concentrate on converting the forecasting problem into a point estimate forecast. Our roadmap for future development includes developing a set of hierarchical models whereby a film will first be classified into a revenue range, as demonstrated in this paper. Subsequently, class-specific models can be employed to forecast more targeted point estimates with corresponding confidence intervals.

Future research directions also include additional experiments in collaboration with one or more major film studios, leveraging additional data sources unavailable to industry outsiders. This would allow for the analysis of the predictive value of granular budgetary and financial data, including specific figures and ratios for actor salaries and special effects, which have previously been studied primarily as qualitative, rather than quantitative, variables. Because production budgets have been widely found to be the single variable with the highest predictive value for box-office revenue forecasting, further study of their constituent parts is warranted. An additional research direction includes algorithmic screenplay analysis. Recent literature has analyzed the text of screenplays with promising results. Because screenplays are available during the pre-production period, they represent information-rich sources of variables with predictive potential. In other research, we have shown DAN2 to be very effective for textual analysis. Application of DAN2 to movie script analysis will be the subject of future investigation as well.

We acknowledge that using any neural network model to forecast box-office revenue faces two primary limitations. The first limitation is the reliance of this approach upon available, sufficient, and reliable data, which is often proprietary. This limitation is magnified when using a multiple class classification method. In this research we developed nine separate models, one for each class, with classes A and B having the fewest in-class records (blockbusters and near-blockbusters, respectively). Although DAN2 performed reasonably well for these two classes, replicating the models with SVM required balancing the datasets in order to assist SVM in producing reasonable results.

Another challenge in using any data driven model, such as an ANN, is the necessity of acquiring sufficient data. Some of the required data for the model, such as advertising expenditures, are proprietary. Publicly available aggregate budget data (i.e. total production budget) does not allow for the analysis of specific budgetary components (e.g. actor wages, special effects budgets, etc.). We also acknowledge that an ANN is characterized as a "black-box" approach. This property, shared by DAN2, does not allow researchers to establish or test for causation in the same manner as is possible with other tools such as regression-based approaches. The only practical option to remedy this limitation is an approach whereby two models are developed, one with and one without the targeted variable. Even in such cases, the impact and size of the effect cannot be statistically explained or validated.

## References


Ainslie, A., Drèze, X., & Zufryden, F. (2005). Modeling movie life cycles and market share. *Marketing Science, 24*(3), 508–517.

Alon, I., Qi, M., & Sadowski, R. (2001). Forecasting aggregate retails sales: A comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services, 8*, 147–156.

Apala, K. R., Jose, M., Motnam, S., Chan, C. C., Liszka, K. J., & de Gregorio, F. (2013). Prediction of movies box office performance using social media. In International conference on advances in social 11 networks analysis and mining (ASONAM). 2013 IEEE/ACM (pp. 1209–1214).

Asur, S., & Huberman, B. A. (2010). *Predicting the future with social media. International conference on web intelligence and intelligent agent technology (WI-IAT), 2010*. IEEE/WIC/ACM (pp. 492–499). IEEE/WIC/ACM.

Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science, 15*(1), 215–227.

Basuroy, S., & Chatterjee, S. (2008). Fast and frequent: Investigating box office revenues of motion picture sequels. *Journal of Business Research, 61*, 798–803.

Basuroy, S., Chatterjee, S., & Ravid, A. (2003). How critical are critical reviews? The box office effects of film critics, star power, & budgets. *Journal of Marketing, 67*, 103–117.

Canterbery, R., & Marvasti, A. (2001). The U.S. motion pictures industry: An empirical approach. *Review of Industrial Organization, 19*, 81–98.

Chen, X., Chen, Y., & Weinberg, C. B. (2013). Learning about movies: the impact of movie release types on the nationwide box office. *Journal of Cultural Economics, 37*(3), 359–386.

Chevalier, J., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*, 345–354.

De Silva, B., & Compton, R. (2014). Prediction of foreign box office revenues based on wikipedia page activity. In *Proceedings of the ACM web science 2014 conference*, Preprint.

De Silva, I. (1998). Consumer selection of motion pictures. In B. Litman (Ed.), *The motion picture mega-industry*. Boston, MA: Allyn & Bacon Publishing Inc.

De Vany, A., & Walls, W. (2004). Motion picture profit, the stable Paretian hypothesis, and the curse of the superstar. *Journal of Economic Dynamics & Control, 28*, 1035–1057.

Delen, D., & Sharda, R. (2010). Predicting the financial success of hollywood movies using an information fusion approach. *Industrial Engineering Journal, 21*(1), 30–37.

Dellarocas, C., Farag, N., & Zhang, M. (2005). Using online ratings as a proxy of word-of-mouth in motion picture revenue forecasting. Working Paper.

Dellarocas, C., Zhang, X., & Awad, N. (2007). Exploring the value of online products reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing, 21*(4), 23–45.

Derrick, F. W., Williams, N. A., & Scott, C. E. (2014). A two-stage proxy variable approach to estimating movie box office receipts. *Journal of Cultural Economics, 38*(2), 173–189.

Desai, M. A., Loeb, G. J., & Veblen, M. F. (2002). The strategy and sources of motion picture finance. Harvard business school background Note, 9-203-007.

Dhar, T., Sun, G., & Weinberg, C. (2012). The long-term box office performance of sequel movies. *Marketing Letters, 23*, 13–29.

Doshi. L. (2010). *Using sentiment and social network analyses to predict opening-movie box-office success* (Master's Thesis). MIT.

Duan, W., Gu, B., & Whinston, A. (2008a). Do online reviews matter? — An empirical investigation of panel data. *Decision Support Systems, 45*, 1007–1016.

Duan, W., Gu, B., & Whinston, A. (2008b). The Dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *Journal of Retailing, 84*(2), 233–242.

Edwards, D., Buckmire, R., & Ortega-Gingrich, J. (2013). A mathematical model of cinematic box-office dynamics with geographic effects. *IMA Journal of Management Mathematics*, 1–25.

Einav, L. (2007). Seasonality in the U.S. motion picture industry. *The RAND Journal of Economics, 38*(1), 127–145.

Einav, L. (2010). Not all rivals look alike: Estimating an equilibrium model of the release date timing game. *Economic Inquiry, 48*(2), 369–390.

Einav, L., & Ravid, S. (2009). Stock market response to changes in movies' opening dates. *Journal of Marketing, 33*, 311–319.

Elberse, A. (2007). The power of stars: Do star actors drive the success of movies? *Journal of Marketing, 71*, 102–120.

Elberse, A., & Anand, B. (2007). The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market. *Information Economics and Policy, 19*, 319–343.

Elberse, A., & Eliashberg, J. (2003). Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Science, 22*(3), 329–354.

Eliashberg, J., Elberse, A., & Leenders, M. (2006). The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing Science, 25*, 638–661.

Eliashberg, J., Hui, K., & Zhang, J. (2007). From storyline to box office: A new approach for green-lighting movie scripts. *Management Science, 53*, 881–893.

Eliashberg, J., Hui, S. K., & Zhang, Z. J. (2014). Assessing box office performance using movie scripts: A kernel-based approach. *IEEE Transactions on Knowledge and Data Engineering, 26*(11), 2639–2648.

Eliashberg, J., Jonker, J., Sawhney, M., & Wierenga, B. (2000). MOVIEMOD: An implementable decision support system for prerelease market evaluation of motion pictures. *Marketing Science, 19*(3), 226–243.

Eliashberg, J., Weinberg, C., & Hui, S. (2008). Decision models for the movie industry. *Handbook of Marketing Decisions, 13*, 437–467.

Fetto, J. (2010, February 20). *American movie-goer consumer report.* Retrieved February 15, 2012, from Experian Marketing Services, Inc: http://www.experian.com/blogs/marketing-forward/2010/02/20/2010-american-movie-goer-consumer-report/.

Foutz, N., & Jank, W. (2007). The wisdom of crowds: Pre-release forecasting via functional shape analysis of the online virtual stock market. In *ACM conference on electronic commerce*. San Diego, California.

Ghiassi, M., & Burnley, C. (2010). Measuring effectiveness of a dynamic artificial neural network algorithm for classification problems. *Expert Systems with Applications, 37*(4), 3118–3128.

Ghiassi, M., & Nangoy, S. (2009). A dynamic artificial neural network model for forecasting nonlinear processes. *Computers & Industrial Engineering, 57*(1), 287–297.

Ghiassi, M., Olschimke, M., Moon, B., & Arnaudo, P. (2012). Automated text classification using a dynamic artificial neural network model. *Expert Systems with Applications, 39*, 10967–10976.

Ghiassi, M., & Saidane, H. (2005). A dynamic architecture for artificial neural network. *Neurocomputing, 63*, 397–413.

Ghiassi, M., Saidane, H., & Zimbra, D. K. (2005). A dynamic artificial neural network model for forecasting time series events. *International Journal of Forecasting, 21*, 241–262.

Gopinath, S., Chintagunta, P., & Venkataraman, S. (2013). Blogs, advertising, and local- market movie box office performance. *Management Science, 59*(12), 2635–2654.

Guardia-Sebaoun, E., Rafrafi, A., Guigue, V., & Gallinari, P. (2013). Cross-media sentiment classification and application to box-office forecasting. In *Proceedings of the 10th conference on open research areas in information retrieval* (pp. 201–208).

Gutierrez-Navratil, F., Fernandez-Blanco, V., Orea, L., & Prieto-Rodrigues, J. (2012). How do your rivals' releasing dates affect your box office? *Journal of Cultural Economics*, 1–14.

Hand, C., & Judge, G. (2012). Searching for the picture: forecasting UK cinema admissions using Google Trends data. *Applied Economics Letters, 19*(11), 1051–1055.

Hanssens, D., Gilbert-Rolfe, J., Merchant, U., & Moroian, V. (2003). Drivers of marketing spending in motion pictures. The Anderson school at UCLA: Applied management research: Special project.

Hennig-Thurau, T., Houston, M., & Heitjans, T. (2009). Conceptualizing and measuring the monetary value of brand extensions: The case of motion pictures. *Journal of Marketing, 73*, 167–183.

Hennig-Thurau, T., Houston, M., & Walsh, G. (2007). Determinants of motion picture box office and profitability: An interrelationship approach. *Review of Motion Pictures, 1*, 65–92.

Jedidi, K., Krider, R., & Weinberg, C. (1998). Clustering at the movies. *Marketing Letters, 9*(4), 393–405.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of the 10th European conference on machine learning*, (pp. 137–142). Chemnitz, Germany.

Joachims, T. (1999). Transductive inference for text classification using support vector machines. In *Proceedings of the sixteenth international conference on machine learning*, (pp. 200–209). Bled, Slovenia.

Joshi, M., Das, D., Gimpel, K., & Smith, N. (2010). Movie reviews and revenues: An experiment in text regression. In *Proceedings of the 2010 annual conference of the North American chapter of the association for computational linguistics*, (pp. 293–296).

Karniouchina, E. (2008). *Empirical essays on marketing issues in motion pictures* (Dissertation). University of Utah, School of Business.

Kim, T., Hong, J., & Koo, H. (2013). Forecasting box-office revenue by considering social network services in the Korean market. *Sains Humanika, 64*(2).

Kim, S., Park, M., & Park, S. (2013). Exploring the effects of online word of mouth and expert reviews on theatrical movies' box office success. *Journal of Media Economics, 26:2*, 98–114.

Krider, R., & Weinberg, C. (1998). Competitive dynamics and the introduction of new products: The motion picture timing game. *Journal of Marketing Research, 35*, 1–15.

Kulkarni, G., Kannan, P., & Moe, W. (2012). Using online search data to forecast new product sales. *Decision Support Systems, 52*, 604–611.

Lewis, D. D. (1997). *Reuters-21578 text categorization test collection.* Retrieved July 7, 2012, from http://kdd.ics.uci.edu/databases/reuters21578/README.txt.

Litman, B. (1983). Predicting success of theatrical movies: An empirical study. *The Journal of Popular Culture, 16*(4), 159–175.

Litman, B., & Ahn, H. (1998). Predicting financial success of motion pictures. In B. Litman (Ed.), *The motion picture mega-industry*. Boston, MA: Allyn & Bacon Publishing.

Litman, B. R., & Kohl, A. (1989). Predicting financial success of motion pictures: The 80's experience. *The Journal of Media Economics, 2*(1), 35–50.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing, 70*, 74–89.

Liu, Y., Chen, Y., Lusch, R., Chen, H., Zimbra, D., & Zeng, S. (2010). User-generated content on social media: predicting new product market success from online word-of-mouth. In *IEEE: Intelligent systems; special trends & controversies section on market and business intelligence 2.0*, January–February.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

Marr, M., & Orwall, B. (2004). Summer film fight: Again, big-budget movies gird for a box-office brawl; studios' darwinian exercise. *Wall Street Journal*, B1 [April 26].

Marshall, P., Dockendorff, M., & Ibanez, S. (2013). A forecasting system for movie attendance. *Journal of Business Research, 66*(10), 1800–1806.

McKenzie, J. (2012). The economics of movies: A literature survey. *Journal of Economic Surveys, 26*(1), 42–70.

McKenzie, J. (2013). Predicting box offing with and without markets: Do internet users know anything? *Information Economics and Policy, 25*, 70–80.

Mestyán, M., Yasseri, T., & Kertész, J. (2012a). Early prediction of movie box office success based on wikipedia activity big data. *Physics and Society* [Retrieved from http://arxiv.org/abs/1211.0970].

Mestyán, M., Yasseri, T., & Kertész, J. (2012b). Early prediction of movie box office success based on wikipedia activity big data. *PLoS ONE, 8*(8), e71226.

Moon, S., Bergey, P., & Iacobucci, D. (2010). Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. *Journal of Marketing, 74*, 108–121.

MPAA. (2004). *US entertainment industry: 2004 MPAA market statistics.* Motion Picture Association Worldwide Market Research. Retrieved from http://www.immagic.com/eLibrary/ARCHIVES/GENERAL/MPAA_US/M050309M.pdf.

MPAA. (2007). *Entertainment industry market statistics.* Motion Picture Association of America. Retrieved from http://www.immagic.com/eLibrary/ARCHIVES/GENERAL/MPAA_US/M080925E.pdf.

MPAA. (2012). *Theatrical market statistics.* Motion Picture Association of America. Retrieved from http://www.mpaa.org/resources/3037b7a4-58a2-4109-8012-58fca3abdf1b.pdf.

Nash Information Services, LLC. (n.d.). *Movie index*. Retrieved January 1, 2014, from The Numbers: http://www.the-numbers.com/movies/.

Neelamegham, R., & Chintagunta, P. (1999). A bayesian model to forecast new product performance in domestic and international markets. *Marketing Science, 18*(2), 115–136.

Nelson, R., & Glotfelty, R. (2012). Movie stars and box office revenues: An empirical analysis. *Journal of Cultural Economics, 36*, 141–166.

Orbach, B., & Einav, L. (2007). Uniform prices for differentiated goods: The case of the movie-theater industry. *International Review of Law and Economics, 27*, 129–153.

Panaligan, R., & Chen, A. (2013). Quantifying movie magic with google search. Google Whitepaper—Industry Perspectives + User Insights.

Pangarker, N. A., & Smit, E. (2013). The determinants of box office performance in the film industry revisited. *South African Journal of Business Management, 44*(3).

Parimi, R., & Caragea, D. (2013). Pre-release box-office success prediction for motion pictures. In *International conference on machine learning and data mining in pattern recognition (mldm) proceedings*, (pp. 571–585).

Prag, J., & Casavant, J. (1994). An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. *Journal of Cultural Economics, 18*, 217–235.

Radas, S., & Shugan, S. (1998). Seasonal marketing and timing new product introductions. *Journal of Marketing Research, 35*(3), 296–315.

Ravid, S. (1999). Information, blockbusters, and stars: A study of the film industry. *The Journal of Business, 72*(4), 463–492.

Ravid, S., & Basuroy, S. (2004). Managerial objectives, the r-rating puzzle, and the production of violent films. *The Journal of Business, 77*(2), S155–S192.

Rennhoff, A., & Wilbur, K. (2010). The effectiveness of post-release movie advertising. Working Paper. Retrieved from http://ssrn.com/abstract=1106932.

Rui, H., Liu, Y., & Whinston, A. (2010). Chatter matters: How twitter can open the black box of online word-of-mouth. In *Proceedings of the international conference on information systems 2010*, December 12–15, 2010. Saint Louis, Missouri.

Sadikov, E., Parameswaran, A., & Venetis, P. (2009). Blogs as predictors of movie success. In *Proceedings of the third international conference on weblogs and social media* (pp. 304–307).

Sawhney, M., & Eliashberg, J. (1996). A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science, 15*(2), 113–131.

Sharda, R., & Delen, D. (2006). Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications, 30*, 243–254.

Sochay, S. (1994). Predicting the performance of motion pictures. *The Journal of Media Economics, 7*(4), 1–20.

Somlo, B., Rajaram, K., & Ahmadi, R. (2011). Distribution planning to optimize profits in the motion picture industry. *Production and Operations Management, 20*, 618–636.

Song, J. W., & Han, S. J. (2013). Predicting Gross Box Office Revenue for Domestic Films. *CSAM (Communications for Statistical Applications and Methods), 20*(4), 301–309.

Spann, M., & Skiera, B. (2003). Internet-based virtual stock markets for business forecasting. *Management Science, 49*, 1310–1326.

Terry, N., Butler, M., & De'Armond, D. (2005). The determinants of domestic box office performance in the motion picture industry. *Southwestern Economic Review, 32*, 137–148.

Terry, N., Cooley, J., & Zachary, M. (2009). The determinants of foreign box office revenue for English language movies. *Journal of International Business and Cultural Studies, 2*(1).

United States v. Paramount Pictures, Inc., 79 (U.S. Supreme Court 1948). Retrieved from https://supreme.justia.com/cases/federal/us/334/131/case.html.

Vogel, H. L. (2001). *Entertainment industry economics*. Cambridge, New York: Cambridge University Press.

Walls, W. (2005). Modeling movie success when "nobody knows anything": Conditional stable-distribution analysis of film returns. *Journal of Cultural Economics, 29*, 177–190.

Wong, F., Sen, S., & Chiang, M. (2012). Why watching movie tweets won't tell the whole story? Workshop on online social networks (WOSN), August 17, 2012. Helsinki, Finland.

Zhang, L., Luo, J., & Yang, S. (2009). Forecasting box office revenue of movies with bp neural network. *Expert Systems with Applications, 36*, 6580–6587.

Zhang, W., & Skiena, S. (2009). Improving movie gross prediction through news analysis. In *International joint conference on web intelligence and intelligent agent technology* (pp. 301–304). IEEE/WIC/ACM.

Zufryden, F. (1996). Linking advertising to box office performance of new film releases – A marketing planning model. *Journal of Advertising Research*, 29–41 (July–August).

Zufryden, F. (2000). New film website promotion and box-office performance. *Journal of Advertising Research, 40*, 55–64.