

پیش بینی فروش فیلم های سینمایی در ایران با استفاده از شبکه های عصبی و مقایسه آن با روش های آماری

عباس طلوعی اشلقی، دانشیار گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه آزاد اسلامی واحد علوم و تحقیقات

toloie@gmail.com

عباس سقایی، دانشیار، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات
سارا خورسندی کریمی، دانش آموخته کارشناسی ارشد مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات

چکیده

پیش بینی فروش فیلم قبل از شروع اکران عمومی آن به دلیل عدم قطعیت عکس العمل مخاطبین، چالشی بزرگ و مساله ای جذاب و وسوسه انگیز و در عین حال، پیچیده و دشوار است. موفقیت یک فیلم سینمایی، در گرو فروش مناسب و پوشش هزینه های ساخت و سود دهی، دانسته می شود. تا آنجا که به کار بردن عبارت صنعت سینما که در درون خود نگرش و پیش فرض های عوامل موفقیت یک فعالیت صنعتی را نهفته دارد، به جای هنر سینما، عمومیت دارد. همچنین، بازگشت سرمایه ی صرف شده برای ساخت فیلم سینمایی، تنها وقتی که تماشاگر سینما به تماشای آن فیلم بنشینند، ممکن خواهد بود. می توان با شناخت عوامل تاثیر گذار در فروش فیلم، متغیر هایی تعریف کرد و با شناخت و آموزش آن ها به شبکه ی عصبی، با ارائه ی مدلی، میزان فروش فیلم را پیش بینی نمود. در این صورت، ابزار کمکی بسیار قدرتمندی در اختیار تولید کنندگان فیلم، به منظور تضمین وجود حاشیه ی امنی برای سرمایه گذاری انجام شده، قرار می گیرد. در این مقاله، انتخاب متغیر های ورودی، بر مبنای ویژگی ها و مقتضیات صنعت فیلم سازی در کشور ایران و با استفاده از روش های آماری و عقاید صاحبان نظر در این حیطه، صورت گرفته است. پس از تبدیل مساله ی این مطالعه، از مساله ی پیش بینی، به یک مساله طبقه بندی شده، با استفاده از شبکه ی عصبی MLP با چند متغیر ورودی و یک متغیر خروجی، مدل پیش بینی، ساخته شده است. همین مساله با بهره گیری از روش آماری رگرسیون چند متغیره نیز حل شده است. در انتها، روش Fold-Cross-Validation ده گانه، برای اندازه گیری میزان کارایی مدل پیش بینی به کار رفته است. این کارایی، با بررسی موفقیت مدل، در پیش بینی فروش فیلم، در رده بندی صحیح خود، و در یک رده بندی بالا و پایین رده بندی صحیح خود، سنجیده شده است. مقایسه نتایج نشان می دهد که شبکه های عصبی، عملکرد بهتری نسبت به روش های آماری برای حل این مساله پیش بینی داشته اند.

واژگان کلیدی: شبکه ی عصبی، رگرسیون چند متغیره، مساله ی پیش بینی، مساله طبقه بندی شده، پیش بینی، فروش فیلم سینمایی.

مقدمه

شبکه عصبی مصنوعی، ایده ای است برای پردازش اطلاعات که از سیستم های عصبی موجودات زنده الهام گرفته است. با اینکه از ساخت اولین سلول عصبی در نیمه قرن بیستم میلادی زمان زیادی می گذرد، ولی تا دهه اخیر، استفاده از ساختار خاص حل مساله توسط شبکه عصبی، عمومیت نیافته بود. قابلیت هیجان انگیز این ایده در استنتاج معانی مبهم و پیچیده، و در نتیجه، درک و شناسایی الگوها و حل مساله، این ابزار را به روشی محبوب که در تحقیقات به کار برده می شود تبدیل کرده است.

بخشی مهم از تحقیقات دانشگاهی متمرکز بر فهم الگوهایی برای پیش بینی رخداد های آتی است. با درک این ساختارها می توان در حوزه های مورد بحث، سیاست گذاری، تصمیم سازی و تحلیل صحیح ارائه داد که این خود در نهایت به افزایش شناخت نوع بشر از جهان پیرامون آن منجر خواهد شد. حوزه های نظری، دانش های جدید، اقتصاد، تجارت و بانکداری مورد توجه محققین شبکه ی عصبی مصنوعی در کشور ایران بوده اند. اما صنعتی به مانند صنعت سینما به دلیل صعوبت ذاتی، عرفی و طبع ویژه خود، از این توجه دور مانده است. علیرغم بررسی صورت گرفته، فعالیت های آکادمیک ثبت شده ای در خصوص پیش بینی میزان فروش فیلم های سینمایی و عوامل موثر بر آن، شناسایی نشد. در این زمینه تلاش های جالب توجهی متمرکز بر فیلم های تولید شده در کشور ایالات متحده آمریکا (شاندر، دلن، ۲۰۰۶) و همین طور متمرکز بر فیلم های به نمایش درآمده در کشور چین (لی ژانگ، ۲۰۰۹)، انجام شده است. تلاش های این گروه، صرفا بر تخمین میزان موفقیت فروش، قبل از آغاز به نمایش در آمدن آن متمرکز بوده است. اما گروهی دیگر از محققین، اطلاعات حاصل از فروش روزهای اولیه نمایش را نیز در نتیجه گیری خود دخیل کرده اند. بدین گونه، با در دست بودن اولین باز خوردهای تماشاگران و منتقدان، پیش بینی میزان موفقیت فروش فیلم، نتایج دقیق تری ارائه می دهد (ساونی، الیاشبرگ، ۱۹۹۶).

میزان توان مندی شبکه عصبی در پیش بینی موفقیت فروش فیلم در بازار ایران، نقطه تمرکز تحقیق پیش رو است که در نظر دارد با شناخت عوامل موثر و میزان تاثیر هر کدام، الگوی قابل استنادی را ارائه دهد و به مقایسه با اطلاعات مدون و مکتوب مراجع سینمایی کشور ایران بپردازد.

از لحظه ای که فردی تصمیم به ساخت فیلم با موضوعی خاص می گیرد، تا زمانی که آخرین حلقه از آخرین نسخه ی فیلم در آخرین سانس سالن نمایش فیلم در معرض دید قرار گذاشته می شود، ده ها عامل و متغیر می توان شمرد که بر موفقیت یک اثر تولیدی صحنه می گذارند و یا باعث شکست آن می شوند. وظیفه محققان این است که علاوه بر تعیین نقش هر کدام از عوامل که به صورت عرفی در این میان نقش ایفا می کنند، عوامل خاص حادث از مدیریت دولتی و اعمال قوانین تشویقی و تنبیهی ویژه و مقتضیات خاص نمایش فیلم در کشور ایران را نیز در نظر آورند. نبود نگاه علمی به فرآیند تولید فیلم، در مراحل نوشتن فیلم نامه، انتخاب بازیگر، بازاریابی و اکران مناسب، موجب لطمه پذیری و افزایش ریسک سرمایه گذاری در صنعت فیلم سازی کشور ایران شده است. نتیجه بلافصل آن، استفاده از راه حل های تکراری و نخ نما در روند تولید فیلم است که محصول خروجی آن ها، عامه پسند، فیلم فارسی و سطحی نگر نامیده میشوند و هویتی مستقل ندارند. ولی از منظر بازگشت سرمایه و حصول سود مناسب، به اطراف قضیه، قابل اطمینان هستند. در کنار آن توجه به ظرایف حاکم بر روند بازاریابی و زمان بندی اکران با توجه به موضوع فیلم، و انتخاب بازیگران ستاره و مورد توجه جامعه با لحاظ نمودن کارکرد سینما در ایران که بعد سرگرمی دارد، از دلایل موفقیت در نظر گرفته شده اند. میزان صحت این نگاه پس از مقایسه ی نتایج به دست آمده و میزان تاثیر گذاری هر یک، با نتایج واقعی در دست، روشن می شود.

ایجاد پشتوانه فکری و علمی برای تصمیم سازان این عرصه، از وظایف اهل علم است. بررسی ها، فعالیتی در زمینه ی تحلیل عوامل سازنده ی شکست و یا موفقیت در ساختن فیلم در کشور ایران را نشان نمی دهد. مقاله هایی در محیط های علمی خارج از کشور ایران یافت شده که در بخش قبل به آن پرداخته شده است.

بعد مهمی از فرآیند تولید و عرضه فیلم، مقطعی است که سازندگان، قصد معرفی اثر ساخته شده خود به تماشاگران بالقوه را دارند. در این تحقیق، پیش فرض مقرر، این است که، همه فیلم های بررسی شده به نحو یکسان و همگون از امکان تبلیغات موثر و کارا، بهره مند شده اند.

تا آغاز جنگ جهانی اول، رشد سینما متمرکز در اروپا بود و از این زمان، ایالات متحده آمریکا که به واسطه وجود اقیانوس آتلانتیک شمالی از گزند تبعات جنگ

مصون ماده بود، جای سلف خود را گرفت. از آغاز ده ۱۹۲۰ میلادی تا انتهای جنگ جهانی دوم در سال ۱۹۴۵، دو رخداد در ابعاد فناوری، به تکمیل این روند کمک کردند. اول تولید فیلم ها با صدا و دوم تولید فیلم به شکل رنگی. از سال ۱۹۱۰ که "گریفیث" برای فیلم برداری فیلمش، به دهکده کوچک و آرام هالیوود در شمال لس آنجلس عزیمت کرد، تا سال ۱۹۸۵ که اولین دوربین های فیلم برداری قابل حمل ویدئویی به بازار عرضه شدند، سینما، شرکت های بزرگ تهیه کننده فیلم و صنعت گران با توان مالی بالا، حاکم بلا منازع عرصه تامین تصویر برای مردمان بودند. اما با سهولت تهیه، عرضه، نمایش و حتی نگهداری تصاویر متحرک، در قالب های بسیار با کیفیت، مانند دی وی های امروزی، لزوم مراجعه به سالن های سینما کمتر احساس می شد. اما صنعتگران عرصه ی سینما که به خوبی خود را همراه با تحولات فناوری تطبیق می دادند، این بار نیز از این تحول عمیق بهره مضاعف بردند و بازار جدیدی برای فروش محصولات خود یافتند و کمی بعد، این تهدید عظیم را به فرصتی عظیم تر بدل کردند. این گونه بود که مردمان جوامع که روزگاری باید در شهر های دور دست، مدت ها به انتظار آمدن نمایش دهنده ی دوره گرد می نشستند تا ساعاتی را با تماشای فیلم، به تمدد و تفریح بگذرانند، دیگر ناچار نیستند تنها در انتهای هفته و روز های تعطیل به سالن های سینما بروند، در هر زمان که اراده کنند، می توانند از مشاهده ی فیلم مورد علاقه خود، لذت ببرند. با بررسی دقیق تر این واقعیت عیان می شود که از میان پنجاه فیلم پرفروش از تاریخ ایجاد این هنر - صنعت، چهل فیلم پس از آغاز هزاره ی سوم میلادی تولید و عرضه شده اند. فارغ از تحولات عظیم سیاسی و فرهنگی و اجتماعی که تولید و عرضه ی تصاویر متحرک بر جوامع گذاشته، که پرداختن به آن از حوصله ی این مقاله فراتر است، چرخه ی اقتصادی این صنعت و منافع حاصل از آن، بسیار برای مردمان اغواکننده است. با توجه به اطلاعاتی که پس از این خواهد آمد می توان پذیرفت که اکنون، سینما به یکی از سود آورترین فعالیت های بشر تبدیل شده است. وجه سرگرمی آن به گونه ای با وجه اقتصادی آن گره خورده است که تهیه کنندگان، به تولید دنباله هایی برای فیلم های موفق پرداخته اند. این گسترش فقط مختص ایالات متحده آمریکا یا اروپا نیست. بلکه کشور هندوستان با تولید سالانه بیش از هزار فیلم سینمایی، هنگ کنگ، چین و دیگر کشور ها با درک سودآوری این صنعت و تاثیر آن در سرگرم سازی، آموزش و تربیت بشر، سرمایه

گذاری های کلانی در حوزه تولید فیلم، تربیت بازیگر و عوامل فیلم سازی، نوشتن فیلم نامه، مالکیت سالن های نمایش، شبکه های توزیع، تبلیغات و تکنولوژی های مرتبط انجام داده اند.

روی دیگر سکه، شکست های بزرگی است که فیلم های تولید شده به بار آورده اند. فیلم هایی که از همان فناوری ها، فیلم نامه نویس ها، بازیگران، توزیع کنندگان و نمایش دهندگانی بهره برده اند، که فیلم های پر فروش رکورد شکن. پس عامل یا عواملی که موجب موفقیت فیلم می شود کدام است؟ سرمایه گذار فیلم از اولین لحظه تا به انتها، این دغدغه را با خود به همراه دارد. باید توجه داشت، به دون مشخص بودن ریسک سرمایه گذار از میزان بازگشت سرمایه، او هرگز حاضر به فعالیت در این صنعت نخواهد بود. امکان پیش بینی موفقیت یک فیلم در فروش، علاوه بر منافع مادی که برای عوامل تهیه آن دارد، این توانایی را به سرمایه گذاران می دهد تا با خاطری آسوده تر به آزمون مرزهای جدید و مقابله با چالش های نو و کمک به رشد ایده های متهورانه تر بپردازند. روش های گوناگون علمی، آماری و تجربی در تمامی این سال ها به کار گرفته شده است تا دغدغه از فروش فیلم را به فرصتی حتمی بدل کند و در افق پیش رو، پیش رفت این هنر - صنعت را تضمین نماید. فعالیت های رو به گسترش محققین در این باره با استقبال ارباب این صنعت مواجه می شود که در گفتار ابتدا و ادامه ی مباحث به آن پرداخته می شود.

جامعه متخصص و ذی صلاح در همکاری با دولت مردان مسوول، به تلاش برای عمق بخشیدن با تاثیرات فرهنگی صنعت سینما مشغول هستند و فعالیت های محسوس و ملموسی در این باره انجام داده اند. اما به نظر می رسد در حوزه ی اقتصادی کمتر تنوعی در نحوه ی فعالیت اربابان صنعت سینما وجود داشته باشد. فرمول ساده ی ستاره سازی و تولید فیلم های عامه پسند با تضمین بازگشت سرمایه و حصول منافع مادی سرشار، مکررا به کار برده می شود و تمایلی به نوآوری و تجربه در فضاهای جدید یا درنوردیدن مرزهای فتح نشده وجود ندارد. هدف، طبعا کاستن از ریسک شکست در این نوع سرمایه گذاری اقتصادی است. به عبارت دیگر، هدف، یافتن متغیر های تاثیر گذار و قابل اطمینان و نیز یافتن روش مناسب برای پیش بینی فروش فیلم ها در اکران عمومی سینمای ایران است. در این تحقیق فرآیند دقیقی به شرح ذیل برای حصول نتیجه، طراحی و اعمال شده اند:

- جمع آوری داده های خام، به روش "داده های از پیش تعیین شده"، از منابع و مراجع سینمای ایران.
- طبقه بندی داده های خام در جدول پایه، برای تحلیل های آتی.
- شناسایی عوامل موثر و متغیر ها، با استفاده از نظرات خبرگان صنعت تولید فیلم ایران، مراجعه به تحلیل های مکتوب و سیاست گذاری های متولیان اداره این صنعت.
- وزن دهی به متغیر ها با استفاده از نظرات خبرگان صنعت تولید فیلم ایران و مقایسه آن با مکتوبات موجود.
- انتخاب مدل با مطالعه ی روش های گوناگون متاهیوربستیکی و آماری.
- وارد کردن داده ها به مدل.
- ثبت خروجی ها و بررسی کارایی آن ها.
- اعتبار سنجی مدل.
- تحلیل و نتیجه گیری.

شکل "داده ها" و طبقه بندی فیلم ها

اولین اقدام، ایجاد پایگاه داده ها است. پس از بررسی و جستجوی بسیار برای استفاده از یک بایگانی جامع از اطلاعات فیلم های به نمایش درآمده، مشخص گردید به دلیل عدم دسترسی به چنین مجموعه ای، باید از کنار هم قرار دادن اطلاعات ثبت شده ی پراکنده، از چندین ماخذ، و پس از اطمینان از صحت و سقم آن ها، مبادرت به ایجاد پایگاه داده ها کرد. از اطلاعات منتشر شده توسط معاونت سینمایی وزارت فرهنگ و ارشاد اسلامی، سازمان سینمایی سوره، موسسه فارابی، اطلاعات عمومی قابل استحصال از اینترنت و اطلاعات مکتوب در مراجع سینمایی، در کنار بایگانی نشریات خبری استفاده شد و پس از مقایسه و رفع موارد ابهام و تناقض و نقص، در نهایت مجموعه ای از اطلاعات ۲۷۴ فیلم بلند سینمایی که در مدت ۶ سال، از ابتدای سال ۱۳۸۳ تا انتهای سال ۱۳۸۸، در سینما های پایتخت، به نمایش عمومی درآمده اند، استخراج و جمع آوری شد. متأسفانه تلاش های انجام شده برای یافتن اطلاعات پیوسته از فیلم های اکران شده پیش از این تاریخ، به سرانجام نرسید. در این مجموعه، نام فیلم، نام کارگردان، نام تا ده بازیگر فیلم، زمان شروع نمایش عمومی، تعداد سینما های نمایش دهنده فیلم، تعداد فیلم هایی که در همان ماه غیر از این فیلم اکران شان آغاز شده، تعداد روزهایی که فیلم روی پرده بوده، تعداد جوایز اصلی، فرعی و ویژه ی

جشنواره فجر، اهدا شده به فیلم، ژانر یا گونه ی فیلم و در نهایت میزان فروش فیلم از محل نمایش عمومی در سینماهای پایتخت، برای هر یک از این ۲۷۴ فیلم، به شکل منظم، مطابق نمونه ی جدول شماره ۱، ثبت شد.

جدول ۱: نمونه ی اطلاعات ثبت شده ی فیلم ها

نام فیلم	کارگردان	بازیگران	زمان اکران	نمایش دهنده تعداد سینماهای	ژانر گونه	تعداد جواز اصلی، فرعی و ویژه جشنواره فجر، اهدا شده به فیلم	غیر از این فیلم اکران شان آغاز شده تعداد فیلمهایی که در همان ماه	تعداد روزهایی که فیلم روی پرده بود	فروش (ریال)
مهمان ما	داریوش مهرجویی	گلاب آدینه، پارسا پیروزفر، امین حیایی، حسن پورشیرازی، زاله علوی، نسرين مقانلو، فریده سپاه منصور	۸۳/۵/۱	۲۰	اجتماعی	۱۰۰۰	۲	۷۰	۴۷۱۳۴۳۱۵۰۰

با اعمال کردن تغییرات شاخص بهای کالاها و خدمات مصرفی در این بازه ی زمانی ۶ ساله، میزان فروش فیلم های سینمایی، تعدیل شده اند. ضرایب تورم با استفاده از اطلاعات رسمی پایگاه اطلاع رسانی بانک مرکزی ایران استخراج شده و به کار رفته است. این ضرایب بر اساس جدول شماره ۲، عبارتند از:

جدول ۲: ضریب تورم اعمال شده بر میزان فروش فیلم ها

سال	۱۳۸۴	۱۳۸۵	۱۳۸۶	۱۳۸۷	۱۳۸۸
ضریب تورم (درصد)	۱۰/۴	۱۱/۹	۱۸/۴	۲۵/۴	۲۵/۴

ضریب تورم برای سال ۱۳۸۳ شمسی به عنوان سال مبنای این تحقیق، طبعاً در نظر گرفته نشد.

همان طور که در قبل ذکر آن رفت، پس از تبدیل مساله ی این مطالعه، از مساله پیش بینی به یک مساله ی طبقه بندی شده، برای فهم بهتر مساله توسط تحلیل گر، فروش فیلم در ۱۲ طبقه، رتبه بندی شده است. به این معنی که در مجموعه داده ها،

به جای رقم فروش فیلم از محل اکران عمومی، یک عدد بین ۱ و ۱۲ به آن اطلاق شد. این ترتیب را در جدول شماره ۳ ملاحظه نمایید:

جدول ۳: طبقه بندی فروش فیلم

شماره طبقه فروش	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
فروش بیشتر از (میلیون ریال)	-	۱۰۰	۲۵۰	۵۰۰	۱۰۰۰	۱۵۰۰	۲۰۰۰	۳۰۰۰	۴۰۰۰	۵۰۰۰	۷۰۰۰	۱۰۰۰۰
فروش کمتر از (میلیون ریال)	۱۰۰	۲۵۰	۵۰۰	۱۰۰۰	۱۵۰۰	۲۰۰۰	۳۰۰۰	۴۰۰۰	۵۰۰۰	۷۰۰۰	۱۰۰۰۰	-

بررسی ۲۷۴ فیلم در بازه ۶ ساله سال ۱۳۸۳ تا سال ۱۳۸۸ که نتیجه ی آن در جدول شماره ۴ منعکس شده، نشان می دهد که علیرغم تمام تمهیدات اندیشه شده توسط صاحبان حقوقی یک فیلم سینمایی، در تمامی این سال ها فیلم های با فروش بسیار کم وجود دارند که نشان از عدم توانایی در پیش بینی صحیح عکس العمل مخاطبین بالقوه فیلم دارد. از آن سو، به دلیل عدم وجود اطلاعات شفاف از هزینه های تولید فیلم های سینمایی، در خصوص سودآور بودن محصول تولید شده نمی توان اظهار نظر کرد. طبعاً از این بابت که فیلمی که هزینه های هنگفتی در هنگام تولید داشته، اگر چه فروش بسیار بالایی داشته باشد، در مقایسه با فیلمی کم خرج با همین میزان فروش ممکن است، به سود کمتری رسیده باشد.

جدول ۴: توزیع طبقه بندی فروش فیلم ها به تفکیک زمان نمایش عمومی

طبقه فروش	۱۳۸۳	۱۳۸۴	۱۳۸۵	۱۳۸۶	۱۳۸۷	۱۳۸۸
۱	۴	۵	۲	۲	۴	۹
۲	۲	۴	۳	۵	۳	۹
۳	۴	۴	۶	۷	۷	۷
۴	۱۰	۳	۲	۷	۱۰	۵
۵	۵	۱	۴	۴	۱	۴
۶	۱	۲	۳	۴	۳	۰

۵	۳	۳	۳	۵	۴	۷
۴	۴	۵	۲	۳	۲	۸
۱	۱	۲	۳	۵	۲	۹
۴	۵	۶	۳	۳	۱	۱۰
۷	۲	۰	۸	۴	۳	۱۱
۱	۴	۲	۲	۱	۵	۱۲

شناسایی متغیر ها

در این تحقیق از دو گروه داده ها استفاده شده است. در گروه اول داده ها، رو وزن دهی به تک تک متغیر ها بر اساس نظر خبرگان انجام شده است. در این مسیر از متغیر های کمی برای آموزش مدل شبکه عصبی و استفاده از روش آماری، بهره گرفته می شود. در گروه دوم داده ها، متغیر کارگردان که متغیر کیفی است با روش مرسوم فازی به متغیر کمی تبدیل شده است. متغیر های بازیگران، زمان اکران و ژانر یا گونه متغیر های کیفی در نظر آورده شده اند. این متغیر ها به صورت ارقام صفر و ۱ در شبکه ی عصبی و روش آماری استفاده شده اند. متغیر های تعداد فیلم های روی پرده در زمان اکران (رقبا)، تعداد روزهایی که فیلم روی اکران بوده است، تعداد جوایز اهدا شده به فیلم و تعداد سینماهای نمایش دهنده ی فیلم به دلیل ماهیتشان همچنان متغیر هایی کمی باقی مانده اند.

متغیر اول، کارگردان: نقش کارگردان، به عنوان مغز متفکر و موتور محرک تولید فیلم سینمایی، بی چون و چرا پذیرفته شده است. کیفیت محصول به دست آمده در هر دو وجه موفقیت و استقبال رسانه ای از دید منتقدین و تماشاگران و نیز شکست در این دو میدان، در ابتدا به پای کارگردان نوشته می شود که در آینده ی شغلی وی تاثیر گذار خواهد بود. بنابر این کارگردان در تعامل با تهیه کننده به عنوان بازوی سرمایه گذاری در فیلم، تلاش می کند تا در تمام شئون فنی و هنری مورد نیاز در مراحل پیش تولید، تولید و پس از آن، از بهترین نیرو ها استفاده کند. او با برقراری ارتباط با مخاطب می تواند تماشاگران وفادار خود را سالیان متمادی به تماشای محصول تولیدی خود بنشانند. در نتیجه، تا آنجا که بر دوش کارگردان است، فروش فیلم به هنگام اکران عمومی، تضمین می شود. "مسعود کیمیایی" مصداق روشنی از

این مطلب است. در بازه ی زمانی ۶ ساله ی ۱۳۸۳ تا ۱۳۸۸، وی چهار فیلم " سربازهای جمعه"، " حکم"، " رئیس" و " محاکمه در خیابان" را به نمایش درآورده است که میزان فروش هر یک به ترتیب در طبقه بندی های ۸، ۹، ۹ و ۱۰ قرار گرفته است. مطابق جدول شماره ۵، پس از این که هر کارگردان در هر کدام از ۵ صفت معرفی شده با لحاظ نمودن نظرات اخذ شده از نخبگان و صاحبان نظر در سینما، قرار گرفت، وزن موثر هر یک نرمال سازی شد. ناگفته پیداست که در مجموعه ی داده ها، هر کارگردان می تواند حائز تعدادی از این شرایط باشد. به نمونه ی جدول شماره ۶ توجه کنید.

جدول ۵: وزن اختصاص داده شده به کارگردان

ناشناس	کم کلر	هنری ساز	پیشکسوت	تجاری ساز
۰/۰۱	۰/۰۴	۰/۱۵	۰/۲	۰/۶

جدول شماره ۶: نمونه ای از صفات تعلق گرفته به کارگردان که در مجموعه ی

داده ها ثبت شده است

کارگردان	تجاری ساز	پیشکسوت	هنری ساز	کم کلر	ناشناس
بهمن فرمان آرا	۰	۱	۱	۰	۰

متغیر دوم، بازیگران: وجه سرگرم کنندگی سینما که به فیلم سازی صفت صنعت را افزود، آنچنان برجسته است که تصور اینکه فیلمی متشکل از نخبه ترین های پشت صحنه، به همراه ماهر ترین بازیگران که به بهترین وجه ایفای نقش کرده اند، ولی متاسفانه صورت های زشتی دارند، از نبرد فتح گیشه، موفق بیرون آید، محال به نظر می رسد. تماشاگر که نوعاً کمال جو و زیبا دوست است، به دنبال آرامش دیداری است که با حضور بازیگران موقر، موجه، توانمند و زیبا در صورت و سیرت، کسب می کند. پس از در نظر گرفتن ۵ صفت برای بازیگران که بر مبنای جمع بندی نظرات متخصصین، به دست آمد، در گروه اول داده ها، مطابق جدول شماره ۷، به روش نرمال سازی، اوزانی به هر کدام اختصاص داده شد. به این ترتیب، متغیر مورد بحث کمی است. در گروه دوم داده ها، تقسیم بندی کیفی ترتیبی بر اساس جدول

شماره ۸، برای این ۵ صفت لحاظ شد. هر کدام از ۱۰ بازیگر متعلق به یک فیلم که در مجموعه ی داده ها ثبت شدند، در ذیل یکی از ۵ صفت قرار گرفتند و مجموع تعداد ایشان، برای آموزش مدل شبکه ی عصبی و استفاده در روش آماری به کار رفت. جدول شماره ۷: وزن اختصاص داده شده به تعداد فیلم های اکران شده، در گروه

اول داده ها

ستاره	معروف	میانسال	پیشکوت	نا آشنا
۰/۴	۰/۳	۰/۲	۰/۱	۰

جدول ۸: رده بندی بازیگران در گروه دوم داده ها

ستاره یا معروف	میانسال یا پیشکوت	نا آشنا
قوی	متوسط	ضعیف

متغیر سوم، ژانر یا گونه ی فیلم: ژانر یا گونه ی فیلم، حاصل نوع فیلم نامه و بیانگر جنس محصول تولید شده توسط سازندگان فیلم است. عبارتی مانند فیلم "پرمخاطب" در درون خود بیانگر این واقعیت است که سازنده ی فیلم با علم به وجه سرگرمی سازی قوی فیلم، جوانب جذب تماشاچی را رعایت کرده است. توجه به این نکته لازم است که ژانر یا گونه ی فیلم، نشانگر کیفیات فنی و هنری فیلم نیست و نمی توان ارزش بیشتری برای فیلمی قائل شد، صرفاً به این خاطر که ذیل ژانر یا گونه ی هنری قرار می گیرد. طبعاً عکس این مطلب نیز صادق است. اختصاص یک ژانر به هر یک از ۲۷۴ فیلم پس از طی روند پیچیده ای ممکن شد. نکته اصلی اینجاست که در موارد متعددی مرز واضحی میان انواع ژانر وجود ندارد. کیفیات فیلم تولید شده به گونه ای است که اختصاص یک عنوان به یک محصول، تنها از عهده ی صاحبان نظر بر می آید. بر اساس مقتضیات سینمای ایران، ۱۴ گونه ی فیلم شناسایی شد و در مجموعه ی داده ها به هر فیلم اطلاق شد. جمع بندی نظر خبرگان امر، در گروه اول داده ها، به شکل متغیر کمی پیوسته در جدول شماره ۹، وزن نرمال سازی شده ای است که به هر گونه فیلم اعطا شد. در گروه دوم داده ها، مطابق جدول شماره ۱۰، ۱۴ گونه ی فیلم در ۳ سطح کیفی پیوسته تقسیم شدند.

جدول ۹: وزن اختصاص داده شده به ژانر یا گونه فیلم در گروه اول داده ها

فیلم فارسی	کمدی	عاطفی	خانوادگی	اجتماعی	جنایی	اکشن
۰/۱۵	۰/۱۵	۰/۱۱	۰/۱	۰/۰۹	۰/۰۸	۰/۰۸
دفاع مقدس	ترازدهی	هنری	کودک	تاریخی	مستند	دینی
۰/۰۸	۰/۰۵	۰/۰۴	۰/۰۲	۰/۰۲	۰/۰۲	۰/۰۲

جدول ۱۰: وزن اختصاص داده شده به ژانر یا گونه فیلم در گروه دوم داده ها

ضعیف	متوسط	خوب
تاریخی هنری کودک مستند دینی	جنایی اکشن دفاع مقدس ترازدهی	اجتماعی خانوادگی عاطفی کمدی فیلمفارسی

متغیر چهارم، تعداد روزهایی که فیلم در حال نمایش بوده است: تصور اینکه در زمان اکران یک فیلم، سالن های نمایش دهنده ی آن خالی از تماشاگر باشند، کمی دور از ذهن است. از این رو با گذر زمان، به میزان فروش فیلم افزوده خواهد شد. این متغیر رابطه مستقیم با فروش فیلم دارد، ولی به دلیل پیش گفته، این رابطه خطی نیست و گاهی مشهود است که فیلمی مدتی بسیار روی پرده بوده ولی فروش چندانی نداشته که طبعاً به دلیل ضعف دیگر متغیرهای تأثیرگذار بر فروش فیلم بوده است. شرایط کاملاً عکس نیز در مواردی برقرار است. این متغیر، متغیری کمی از نوع گسسته می باشد .

متغیر پنجم، زمان اکران: زمان اکران با فروش فیلم رابطه مستقیم دارد. در ماه بهمن، به عنوان آغاز سال جدید سینمایی در کشور ایران، به دلیل برگزاری

جشنواره فیلم فجر، فروش فیلم های روی پرده به وضوح کاهش می یابد. در سالیان اخیر کمتر تهیه کننده ای راغب است تا محصول خود را در فاصله ی هفته های میانی دی ماه منتهی به آغاز جشنواره، تا پایان ماه اسفند به نمایش درآورد. نگاهی به تاریخ آغاز اکران فیلم ها در بازه ۶ ساله ی مورد بررسی نشان می دهد، نمایش های آغاز شده در اسفندماه، همگی در روزهای پایانی سال و به عنوان اکران نوروزی به روی پرده رفته اند و تنها از لحاظ تقویمی در ماه اسفند ثبت می شوند. پس از پایان موج اکران نوروزی و استقبال سنتی صورت گرفته از آن، تا آغاز امتحانات مدارس در پایان سال تحصیلی از خرداد ماه، منتهی به تیر، فروش فیلم ها افت کرده و پس از رونق فصل تعطیلات تابستان، در اواسط پاییز نیز دچار کاهش می شود. در آنچه گروه اول داده ها نامیده شد، پس از اختصاص وزن مربوط به هر ماه که همچون گذشته بر مبنای کسب نظر از صاحبان رای در این حوزه تنظیم شده است، وزن ها مطابق جدول شماره ۱۱، نرمال سازی شده اند. متغیر زمان اکران در گروه اول داده ها، از نوع کمی پیوسته است. در آنچه گروه دوم داده ها نامیده شد، دسته بندی ماه ها به گونه ای صورت گرفته است که ماه هایی که از لحاظ انتخاب، همانند یکدیگر هستند در یک دسته قرار گرفته اند. متغیر زمان اکران در گروه دوم داده ها، مطابق جدول شماره ۱۲، از نوع کیفی ترتیبی است.

جدول ۱۱: وزن اختصاص داده شده به هر ماه از سال شمسی در گروه اول داده ها

فروردین و اسفند	اردیبهشت	خرداد	تیر	مرداد	شهریور
۰/۱۸	۰/۱	۰/۰۹	۰/۰۸	۰/۱۱	۰/۰۹
مهر	آبان	آذر	دی	بهمن	
۰/۰۸	۰/۰۹	۰/۱۲	۰/۰۳	۰/۰۳	

جدول ۱۲: وزن اختصاص داده شده به هر ماه از سال شمسی در گروه دوم داده ها

ها

عالی	خوب	پر مخاطره	متوسط	بد
فروردین اسفند	مرداد	خرداد تیر آبان	اردیبهشت شهریور مهر آذر	دی بهمن

متغیر ششم، تعداد سینماهایی که فیلم در آن ها اکران می شود: این متغیر، متغیری کمی گسسته با رابطه مستقیم با خروجی است. کاملاً مشخص است که هر چه تعداد سینماهایی که فیلم در آن ها اکران می شود، بیشتر باشد، میزان خروجی نیز بیشتر می شود. بر اساس منابع استفاده شده در این تحقیق تعداد ۷۹ سینمای فعال به نمایش فیلم در پایتخت مشغول هستند. لازم به ذکر است رابطه ی این متغیر با سود آوری فیلم، رابطه ی مستقیم نیست و تنها در باره ی فروش، می توان چنین اظهار نظری کرد. به عنوان نمونه سینما داران در ایالات متحده ی آمریکا، تولید کنندگان فیلم هستند. بنابر این هزینه های ثابت نمایش یک فیلم در یک نوبت (سانس) خاص، فارغ از اینکه سالن نمایش مملو از تماشاگر باشد و یا خالی، به پای هزینه های فیلم نوشته می شود. ولی حسب اطلاعات موجود، این موضوع مبتلا به شرایط نمایش در کشور ایران نیست، چون به صاحبان سالن های سینما، درصدی از فروش روزانه به عنوان حق الزحمه خدمات ارائه شده، تعلق می گیرد.

متغیر هفتم، تعداد فیلم های روی پرده در زمان اکران (رقبا):

اگر همزمان با فیلمی که روی پرده است، فیلم های هم تراز دیگری در حال اکران باشند، مسلماً فروش همگی آنها تحت تأثیر قرار می گیرد. به این معنا که تماشاگر بالقوه فیلم در یک زمان حق انتخاب بیشتری برای رفتن به سینما خواهد داشت. حال اگر این تماشاگر از نظر بودجه، هزینه ای که باید برای خرید بلیط و رفت و آمد به سالن نمایش دهنده فیلم و یا زمان فراغت، محدودیت داشته باشد، فروش فیلم با زمانی که تنها یک فیلم، تنها و بی رقیب در حال اکران باشد، کاملاً متفاوت است. در بازه ۶ ساله پیش گفته، فیلم "مارمولک" که در سال ۱۳۸۳ به نمایش عمومی درآمد، مصداق عینی چنین شرایطی است. این فیلم در مدت تنها ۲۸ روز، ولی بر پرده ی تمامی سینماهای پایتخت نمایش داده شد، و میزان فروش آن در طبقه بندی ۱۲، قرار گرفت. اطلاعات موجود نشان می دهد، تعداد فیلم هایی که همزمان روی پرده به نمایش درآمده اند، بین ۱ تا ۹ فیلم بوده است. جدول شماره ۱۳ نشانگر اوزان اختصاص داده شده به هر وضعیت اکران است. این متغیر که رابطه ای معکوس با متغیر خروجی فروش دارد، متغیری کمی و پیوسته است.

جدول ۱۳: وزن متغیر تعداد فیلم هایی که همزمان در حال نمایش هستند

۸	۶	۵	۴	۳	۲	۱	۰
۰/۰۲	۰/۰۷	۰/۰۹	۰/۱۲	۰/۱۴	۰/۱۶	۰/۱۹	۰/۲۱

متغیر هشتم، جوایز: فیلم ها به تبعیت از قوانین متولیان قبل یا بعد از اکران در جشنواره های مختلف شرکت می کنند و ممکن است برنده جوایزی شوند و این موضوع باعث استقبال و ترغیب تماشاگران بالقوه برای رفتن به سینما و در نتیجه تأثیر مستقیم در فروش فیلم، می گردد. جشنواره بین المللی فیلم فجر محل رقابت فیلم ها برای کسب جوایز مختلف هستند. سه دسته جوایز اصلی، فرعی و ویژه به شرح جدول شماره ۱۴، به صورت مستمر در این رقابت به فیلم های برتر اهدا می شوند. در سالیان مختلف، گروه برگزار کننده ی جشنواره، جوایز جنبی دیگری را نیز به فیلم های مورد نظر خود اهدا می کنند که در شمارگان این سه دسته جوایز، آورده نشده است. جمع تعداد هر کدام از جوایز دسته های سه گانه در مجموعه ی داده های متعلق به هر فیلم، ثبت شده است. این متغیر که رابطه ای مستقیم با متغیر خروجی فروش دارد، متغیری کمی و گسسته است.

جدول ۱۴: فهرست جوایز اصلی، فرعی و ویژه ی جشنواره ی فیلم فجر

جوایز اصلی جشنواره فجر، اهدا شده به فیلم شامل	جوایز فرعی جشنواره فجر، اهدا شده به فیلم	جوایز ویژه جشنواره ی فجر، اهدا شده به فیلم
فیلم، کارگردان بازیگر نقش اول مرد بازیگر نقش اول زن فیلم نامه	بازیگر نقش مکمل مرد بازیگر نقش مکمل زن فیلمبرداری، تدوین طراحی لباس و صحنه چهره پردازی، صدابرداری صداگذاری، موسیقی جلوه های ویژه	فیلم منتخب تماشاچیان جوایز ویژه موردی

وزن کلی اختصاص داده شده به متغیرها:

با روش های کسب اطلاع پیش گفته شده و با روش نرمال سازی، وزن اختصاص داده شده به متغیرهای گروه اول داده ها و گروه دوم داده ها برای لحاظ نمودن تاثیر کلی هر متغیر در تعیین و پیش بینی میزان فروش فیلم، مطابق جدول ۱۵ در نظر گرفته شده است.

جدول ۱۵: وزن کلی اختصاص داده شده به هر متغیر در گروه اول داده ها و گروه دوم داده ها

جوایز	تعداد فیلم های روی پرده در زمان اکران	تعداد سینما های نمایش دهنده فیلم	زمان اکران	تعداد روزهایی که فیلم روی پرده بود	ژانر	بازیگران	کارگردان
۰/۰۴	۰/۰۸	۰/۳	۰/۰۶	۰/۳۵	۰/۱۲	۰/۱۵	۰/۱

روش پیش بینی

الف) شبکه عصبی:

مدل های شبکه های عصبی برای پیش بینی موفقیت مالی طراحی شده اند و این مدل ها می توانند به عنوان ابزاری قدرتمند برای استودیوها و تهیه کنندگان و سرمایه گزاران فیلم به منظور پیش بینی فروش و موفقیت آن ها بکار روند. بدون هیچ اغمازی، مدل شبکه عصبی ساخته شده با دو لایه پنهان، دقیق ترین خروجی ها را به دست می دهد. در این تحقیق، از شبیه ساز شبکه عصبی نرم افزار MATLAB برای ساخت مدل شبکه عصبی MLP با چند متغیر ورودی برای پیش بینی خروجی رده بندی شده، استفاده شد. گروه اول داده ها، تعداد ۸ متغیر ورودی مستقل و دو لایه ی پنهان برای آموزش مدل استفاده شد. در گروه دوم داده ها، تعداد ۸ متغیر ورودی مستقل و دو لایه پنهان برای آموزش مدل استفاده شد. طبیعت شبکه ی عصبی ایجاب می کرد، آزمون های زیادی برای بهینه کردن خروجی با ایجاد تغییراتی در ۴ گزاره `net.trainParam.goal` `net.trainParam.epochs` `net.trainParam.lr` و `net.trainParam.max_fail` در هر مرحله اجرای کد نوشته شده در نرم افزار MATLAB، انجام گیرد.

ب) روش آماری:

در بررسی روابط آماری فرضیات مختلف تحقیق، متغیرها به صورت دو به دو مورد تحلیل قرار می گیرند. برای مطالعه ی تأثیر همزمان متغیرهای مستقل بر روی متغیر وابسته از روش آماری تحلیل رگرسیون چند متغیره با استفاده از نرم افزار SPSS و تحلیلگر داده ی PASW Statistics کمک گرفته شده است. به صورت سنتی، روش ورود

متغیر ENTER به کار رفت. پس از ورود مجموعه ی داده های گروه اول و گروه دوم، خروجی ها و تحلیل ها به ترتیبی که در بخش های آتی شرح داده می شود ثبت شد. و در نهایت، برای بررسی خروجی نرم افزار، نتایج آزمون تحلیل رگرسیون به مانند "ضریب همبستگی چندگانه"، "ضریب تعیین"، "ضریب تعیین تعدیل یافته"، "خطای معیار"، "تحلیل واریانس" و "سطح معناداری"، در هر مرحله کنترل گردید.

نحوه ثبت خروجی

همان طور که به عنوان نمونه در جدول شماره ۱۶ قابل مشاهده است، در مجموعه داده های گروه اول و دوم و در هر دو روش شبکه ی عصبی و روش آماری، برای مشاهده میزان دقت پیش بینی فروش داده های گروه اول، از تفاضل مقدار واقعی و مقدار پردازش شده استفاده شده است که هر چه حاصل کوچکتر باشد، خطای کمتری را نشان می دهد.

جدول ۱۶: مقایسه فروش واقعی یک نمونه فیلم با خروجی شبکه ی عصبی در گروه اول داده ها

نام فیلم	سال اکران	فروش واقعی رتبه بندی شده	فروش پیش بینی شده	فروش پیش بینی شده گرد شده	اختلاف فروش واقعی و فروش پیش بینی شده
میم مثل مادر	۸۵	۱۲	۱۱/۲۲	۱۱	۱

کارایی شبکه عصبی

همان طور که در قبل ذکر شد، برای بهره بردن از قدرت شبکه عصبی، پیش بینی فروش فیلم، از یک مساله پیش بینی به یک مساله ی طبقه بندی شده تبدیل شد. به این شکل که مدل شبکه ی عصبی ورودی ها را دریافت می کند و خروجی مدل، یعنی فروش فیلم، یکی از بازه های دوازده گانه ی طبقه بندی شده است. درصد نرخ موفقیت برای اندازه گیری کارایی مدل شبکه ی عصبی به کار می رود. این نرخ که "درصد صحت پیش بینی" نامیده می شود، در مسائل طبقه بندی کاربرد بسیار دارد

													پیش بینی

جدول ۱۸: مقایسه فروش واقعی فیلم با فروش پیش بینی شده توسط شبکه عصبی برای گروه دوم داده ها فروش واقعی فیلم

	۱۲	۱۱	۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱	
	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱
	۰	۱	۰	۰	۰	۰	۰	۱	۱	۸	۴	۷	۲
	۰	۰	۰	۰	۱	۰	۰	۳	۱۳	۱۴	۱۷	۱۷	۳
	۰	۰	۱	۰	۱	۰	۱	۱	۱۱	۸	۳	۱	۴
	۰	۰	۱	۰	۰	۴	۳	۴	۵	۳	۰	۰	۵
	۰	۰	۰	۰	۱	۶	۳	۶	۳	۱	۰	۰	۶
	۲	۰	۲	۱	۴	۷	۴	۳	۲	۰	۲	۰	۷
	۲	۰	۲	۵	۳	۳	۱	۰	۲	۰	۰	۰	۸
	۱	۲	۷	۱	۳	۳	۰	۱	۰	۱	۰	۰	۹
	۵	۱۴	۶	۵	۴	۰	۲	۰	۰	۰	۰	۰	۱۰
	۵	۷	۳	۲	۳	۰	۰	۰	۰	۰	۰	۱	۱۱
	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱۲
													درصد صحت پیش بینی در بازه مورد نظر
	۳۳	۸۸	۷۳	۷۹	۵۰	۷۰	۶۹	۵۸	۷۸	۸۶	۸۱	۲۷	

کارایی روش آماری

برای اینکه بتوان مقایسه ی مناسبی میان نتایج حاصل از خروجی پیش بینی شبکه عصبی با روش آماری در دست داشت، کارایی این روش، نسبت مجموع تعداد فیلم هایی است که در روش آماری، فروش آن ها در بازه ی درست و یا با یک بازه اختلاف طبقه بندی شده است، به مجموع فیلم هایی که فروش واقعی آنها در آن بازه قرار دارد. این عدد به صورت درصد بیان می شود. جدول شماره ۱۹ متعلق به گروه اول داده ها و جدول شماره ۲۰ متعلق به گروه دوم داده ها در حالت تفکیک فروش فیلم ها به ۱۲ رده هستند.

جدول ۲۱: سنجش اعتبار مدل پیش بینی به روش Fold-Cross-Validation ده گانه

داده های گروه دوم		داده های گروه اول		تعداد مجموعه های ده گانه
روش آماری	شبکه عصبی	روش آماری	شبکه عصبی	
۵۲/۰۳	۷۸/۷۹	۷۶/۹۷	۷۰/۶۹	۱
۷۲/۹۲	۷۶/۱۴	۶۰/۹۸	۶۳/۶۴	۲
۸۴/۵۵	۴۹/۳۹	۸۵/۵۶	۸۰/۵۶	۳
۹۵/۸۳	۷۰/۸۳	۸۴/۸۵	۸۳/۰۶	۴
۷۲/۱۷	۵۲/۵۰	۷۰/۰۰	۷۸/۱۷	۵
۹۸/۴۸	۴۲/۴۲	۸۰/۳۰	۷۲/۴۲	۶
۹۵/۶۳	۹۴/۰۵	۸۰/۵۶	۸۰/۵۶	۷
۵۲/۷۳	۳۶/۳۶	۶۲/۸۸	۹۵/۴۵	۸
۷۲/۴۹	۸۴/۲۶	۶۳/۳۳	۷۰/۳۷	۹
۴۷/۱۲	۴۷/۵۸	۷۷/۷۳	۶۳/۴۸	۱۰
۷۴/۴۹	۶۳/۲۳	۷۴/۸۲	۷۵/۷۴	میانگین درصد صحت پیش بینی

مطابق نتایج منعکس شده در جدول ۲۱، در شرایط تفکیک فروش فیلم ها به ۱۲ رده با توجه به درصد میانگین بالاتر، عملکرد شبه های عصبی در گروه اول داده ها به نسبت روش آماری بهتر بوده است. توجه به این نکته ضروری است که روش های آماری سنتی ریاضی در جامعه ای از داده ها معادل ۱۰ درصد جامعه اصلی امکان پیش بینی بالایی به دست خواهند داد. از نتایج این مقایسات و اعتبار سنجی مدل مشخص می گردد، مدل شبکه عصبی برای حل این مساله، مدلی معتبر است. با داشتن داده های ورودی برای هر فیلم سینمایی ساخته شده و یا حتی در حال ساخت و قرار دادن آن ها در جای متغیر های ورودی مدل، با تقریب مناسبی می توان موفقیت آن فیلم در گیشه را پیش بینی کرد.

نتیجه گیری

مشکلات قابل ذکری به هنگام جمع آوری اطلاعات برای این تحقیق پدید آمد. بین داده های معاونت سینمایی وزارت فرهنگ و ارشاد اسلامی، سازمان سینمایی سوره، موسسه ی فارابی، اطلاعات عمومی قابل استحصال از اینترنت و اطلاعات مکتوب در مراجع سینمایی، تناقض های پیاپی به چشم خورد و این اطلاعات به صورت کامل نیز موجود نبودند. اخیراً، موسسه ی فارابی، پایگاه ثبت داده های فیلم های به نمایش درآمده را طرح ریزی کرده است که از ابتدای سال ۱۳۸۹، داده ها را ثبت کرده و برای استفاده عموم، در اختیار قرار می دهد. اما در برابر آن، اطلاعات عرضه شده در پایگاه اینترنتی معاونت سینمایی وزارت فرهنگ و ارشاد اسلامی، که فروش های روزانه فیلم های در حال اکران را ارائه میداد، حذف شده است. لازم است مدیریت واحدی، برای یکسان سازی داده های قابل دسترس عموم در بازه ی زمانی سال ۱۳۸۳ تا پایان سال ۱۳۸۸ که در این تحقیق مورد استفاده قرار گرفته است، و همچنین قبل از سال ۱۳۸۳، اقدام کند.

افزایش تعداد متغیر ها و در نظر آوردن عوامل تاثیر گذار دیگر در فروش فیلم به هنگام اکران عمومی، مانند موسیقی فیلم و سازنده ی آن موسیقی، روش های به کار رفته برای تبلیغات قبل از شروع اکران و هم زمان با آن، و قدرت و نفوذ تهیه کننده ی فیلم، تعداد دفعات (سانس) نمایش فیلم در روز در هر سینما، تعداد دفعات نمایش فیلم به طور فوق العاده در انتهای شب و انتهای هفته، تعطیلات و مناسبت های مذهبی و ملی که بحث عدم تطبیق تقویم های شمسی و قمری را در خود نهفته دارد، امکان استفاده صاحبان فیلم از تبلیغات تلویزیونی، به دست آوردن و لحاظ نمودن داده های حاصل از نمایش فیلم ها در شهر های دیگر غیر از پایتخت و حتی ملاحظه ی ایامی که به طور عرفی و شرعی محذوراتی در تفریحات عمومی وجود دارد به مانند سه ماه محرم و صفر و رمضان، به مدل برای ارائه ی پیش بینی صحیح تر کمک خواهد کرد. چالش های بزرگ در مسیر استفاده از این اطلاعات، دستیابی به داده های فوق الذکر و پس از آن قابل سنجش نمودن آن ها است. این امکان بالقوه وجود داشت تا متغیر های دیگری نیز برای محاسبات در نظر آورده شوند که البته نادیده گرفته شدند. توجه به این دلیل عمده لازم به نظر می رسد که اندازه گیری بعضی متغیر ها عملاً ناممکن بود. همان طور که در جای جای این مقاله، عنوان شد، نبود اطلاعات مضبوط

در بازه های زمانی چند ساله، قابل دسترس و ارائه از سوی متولیان امر، هیچ شانسی برای دخیل نمودن متغیر های دیگری مانند "فروش روزانه فیلم در ایام هفته" یا "فروش روزانه فیلم در روز های با بلیط نیم بها" یا "بودجه ساخت فیلم"، باقی نگذاشت. کمترین تاثیر وجود چنین متغیر هایی، افزایش دقت در قضاوت و یادگیری مدل است. تقسیم بندی متغیر ها به اشکال دیگر نیز می تواند به کیفیت یادگیری شبکه عصبی کمک کند و در نتیجه پیش بینی انجام شده توسط مدل دقیق تر انجام گیرد. در این باره مشخصاً می توان به متغیر کارگردان اشاره نمود: در این تحقیق صفاتی که به کارگردان اطلاق شده برگرفته از محصول فعالیت او یعنی فیلم سینمایی است. صفت "تجاری ساز" یا "هنری ساز" مستقیماً به این نکته اشاره دارد. تقسیم بندی دیگری با استفاده از صفات شخصی کارگردان، مانند "سن" یا "تحصیلات آکادمیک مرتبط" و بدون اشاره به جنس فیلم ساخته شده توسط او می تواند بررسی شود. در خصوص اطلاعاتی که با استفاده از نظرات خبرگان این صنعت بدست آمده است، می توان نظرات گروه های متعددی از ایشان در هر تخصص سینمایی، را با استفاده از نظر سنجی های گسترده تر و عمیق تر که طبعاً مستلزم صرف وقت و هزینه و فعالیت جمعی و گروهی است، کسب کرد که موجب دقیق تر شدن در پی ریزی شالوده اطلاعات خواهد شد.

منابع

- عباس طلوعی اشلقی، رضا رادفر، مدل سازی شبکه های عصبی، انتشارات دانشگاه آزاد اسلامی واحد علوم و تحقیقات، ۱۳۸۴
- Predicting box-office success of motion pictures with neural networks Ramesh Sharda, Dursun Delen
Expert Systems with Applications Vol.30 (2006).
- Forecasting box office revenue of movies with BP neural network Li Zhang, Jianhua Luo, Suying Yang
Expert Systems with Applications Vol.36 (2009)
- A parsimonious model for Forecasting gross box-office revenues of motion pictures
Sawhney, M. S., & Eliashberg, J. Marketing Science, Vol.15 (1996).