دوازدهمین کنگره ملی سراسری
فناوریهای نوین در حوزه توسعه پایدار ایران
12th National Congress of
the New Technologies in Sustainable Development of Iran
senaconf.ir

پیش‌بینی محبوبیت فیلم و مخاطب هدف با استفاده از سیستم توصیه‌کننده مبتنی بر محتوا

محمدامین عنایتی[1]، علی بخت همت[2]

[1]موسسه عالی جهاد دانشگاهی استان خوزستان، اهواز Enayati2013@Gmail.com

[2] استادیار، موسسه جهاد دانشگاهی استان خوزستان، bakhthemmat.std@gmail.com

چکیده

فیلم یکی از اجزای جدایی ناپذیر سرگرمی روزمره ماست. صنعت سینما در سراسر جهان یکی از صنایع رو به رشد و قابل توجه بوده که توجه افراد در هر سنی را به خود جلب کرده است. در مطالعات اخیر مشاهده شده که تنها تعداد کمی از فیلم ها به موفقیت می رسند لذا بلاتکلیفی در این بخش فشار زیادی را بر ذینفعان تولید فیلم ایجاد کرده است. فیلمسازان و محققان به طور مداوم احساس می کنند که لازم است سیستم های خبره ای داشته تا احتمال موفقیت فیلم را قبل از تولید آن با دقت معقول پیش بینی نمایند. کارهای پژوهشی زیادی برای پیش بینی محبوبیت فیلم در مرحله پس از تولید انجام شده است. برای کمک به سازنده ف یلم در تخم ین ف یلم مناسب تماشاگران در آینده و ایجاد تغییرات لازم، باید پیش بینی را در مراحل اولیه تولید فیلم انجام داده و اقدامات مقتض ی را در مورد ف یلم انجام داد.

این پژوهش یک سیستم توصیه فیلم (RS) مبتنی بر محتوا (CB) با استفاده از ویژگی‌های اولیه فیلم مانند ژانر، بازیگران، کارگردان، کلمات کلیدی و توضیحات فیلم پیشنهاد کرده است. با استفاده از خروجی RS و رتبه‌بندی فیلم و اطلاعات رای‌گیری فیلم‌های مشابه، مجموعه‌ای از ویژگی‌های جدید ایجاد کرده و یک مدل یادگیری عمیق CNN (DL) را برای ساخت یک سیستم پیش‌بینی محبوبیت فیلم چند کلاسه پیشنهاد کردیم. ما همچنین سیستمی را برای پیش‌بینی محبوبیت فیلم های آتی در ب ین گروه‌های مختلف مخاطب پیشنهاد نمودیم. گروه مخاطب را به چهار گروه سنی خردسال، نوجوان، میانسال و سالمند تقسیم کردیم. در این مطالعه از داده‌های پایگاه داده فیلم اینترنتی (IMDb) که در دسترس عموم قرار گرفته و داده‌های پایگاه داده فیلم (TMDb) استفاده شده است. ما یک مدل طبقه‌بندی چند کلاسه را پیاده‌سازی کرده که منجر به دستابی دقت ۹۶.۸٪ شده که از همه مدل‌های معیار نیز برتری دارد. این مطالعه پتانسیل تجزیه و تحلیل داده های پیش بینی کننده و تجویزی در سیستم های اطلاعاتی را برای حمایت از تصمیمات صنعت برجسته می کند.

واژه‌های کلیدی

سیستم های خبره، مبتنی بر محتوا، سیستم توصیه، یادگیری عمیق، گروه های مخاطب

۱- مقدمه

صنعت سینما در سراسر جهان یک صنعت درآمدزا بوده و از این رو این صنعت دارای گردش مالی چندین میلیارد دلاری است. اکثر مردم با این صنعت در ارتباط هستند و سرمایه گذاری هنگفتی در آن وارد می شود. در سال ۲۰۱۹ کل درآمد باکس آفیس ایالات متحده و کانادا ۱۱.۳۲ میلیارد دلار بوده است [۵۹].

با این حال، تعداد کمی از فیلم های تولید شده به موفقیت دست یافته اند. تولیدکنندگان و محققان فیلم دائماً احساس می‌کنند که داشتن سیستم‌های خبره‌ای ضروریست زیرا که شانس موفقیت فیلم را پیش‌بینی کرده و تولید آن را با دقت مناسب می سنجد. صنعت سینما بسیار گسترده و متنوع است. پارامترهای مختلف زیادی در ایجاد یک فیلم نقش دارند. نمایش موفقیت یا درجه آن در آینده یک فیلم، کار بسیار پیچیده است. کارهای تحقیقاتی [۳۲]، [۳۳]، [۶۰] زیادی برای پیش بینی محبوبیت فیلم انجام شده است. پیش از این، چندین کار در زمینه پیش‌بینی پس از تولید یا پس از انتشار انجام شده است. با این حال، بدلیل اینکه سرمایه گذار قبلاً وجوه خود را به تولید آن اختصاص داده، سودمند نیست. پیش‌بینی‌ای که بلافاصله پس از مشخص شدن بازیگران، کارگردان و خط داستانی انجام شود، به سرمایه‌گذار در تصمیم‌گیری مالی کمک می‌کند.

پس از یک مطالعه دقیق، ما شاهد تحقیقات قابل توجهی در مورد پیش بینی موفقیت فیلم قبل از انتشار رسمی بودیم. پیش‌بینی‌هایی که کمی قبل از [۳۲]، [۵۶] یا پس از [۳۳]، [۶۰]، [۳۴] اکران رسمی (آخرین مرحله در تولید فیلم) انجام می‌شوند ممکن است داده‌های بیشتری برای استفاده و تولید پیش‌بینی دقیق‌تری داشته باشند [۶۶]. با این حال، آنها به طور قابل توجهی برای سرمایه گذاران برای برآورد هر تصمیم حیاتی به تاخیر می افتند. پیش‌بینی در مراحل اولیه (تولید) [۳۹]، [۴۰] سودمندترین راه در موفقیت فیلم است. کار بسیار کمی برای پیش بینی موفقیت فیلم در مراحل اولیه تولید فیلم انجام شده است. دقت پیش‌بینی در مراحل اولیه آثار قبلی، چندان خوب نیست. اکثر پیش بینی ها فقط برای تمرکز بر احتمال موفقیت فیلم در آینده انجام می شود. برخی از آنها مسئله را به یک مسئله باینری (Hit/Flop) و در برخی کارها، آنها را به یک مسئله چند کلاسه طبقه بندی می کنند. سازندگان فیلم شروع به ساخت یک فیلم جدید می کنند در حالی که بیشتر اوقات یک گروه یا گروه مخاطب خاص را هدف قرار می دهند. سن مخاطب یکی از معیارهای ضروری برای مخاطب هدف است [۶۱]. برخی از فیلم ها با هدف قرار دادن گروه مخاطب خردسال ساخته می شوند. برخی از فیلم‌ها مخاطبان نوجوان را هدف قرار می‌دهند، گاهی اوقات مخاطبان میانسال و بزرگسال را هدف قرار می‌دهند و برخی فیلم‌ها برای همه هستند. فرض کنید می‌توانیم پیش‌بینی کنیم که آیا فیلم در آینده در میان کدام گروه مخاطبان شهرت خواهد داشت یا خیر. اگر بتوانیم تأثیرگذاری فیلم در آینده را در بین تمام گروه های سنی در مراحل اولیه تولید فیلم بسنجیم، سازندگان فیلم سود خواهند برد. سپس، فیلم سازان می توانند در صورت نیاز تغییرات لازم را ایجاد کنند. بین پیش‌بینی موفقیت فیلم و پیش‌بینی مخاطب هدف در مراحل اولیه تولید ارتباط معناداری وجود دارد.

در این پژوهش مدلی جهت پیش‌بینی موفقیت فیلم در مراحل اولیه تولید و انجام پیش‌بینی مخاطب هدف فیلم پیشنهاد می شود. هر دو کار فوق با استفاده از سیستم توصیه فیلم CB انجام شده است. کار تحقیقاتی ما را می توان به سه بخش مهم تقسیم کرد. در چارچوب ما، اولین ماژول یک سیستم توصیه فیلم است [۲]. ما تنها پنج ویژگی اساسی فیلم مانند ژانر، بازیگران، کارگردان، کلمات کلیدی و توضیحات فیلم را به عنوان مجموعه ویژگی برای ساختن سیستم توصیه در نظر گرفته‌ایم. همه این ویژگی های اساسی در مراحل اولیه تولید فیلم در دسترس هستند. سیستم پیشنهادی مجموعه‌ای از فیلم‌های مشابه یک فیلم در آینده را ارائه می‌کند. ماژول دوم فیلم های مشابه ماژول اول را می پذیرد، از رتبه بندی فیلم و اطلاعات رای گیری فیلم های مشابه استفاده می کند و یک مجموعه ویژگی جدید ایجاد می کند. در مرحله بعد، ما یک مدل CNN پیشنهاد کرده‌ایم و از مجموعه داده‌های جدید ایجاد شده برای پیش‌بینی محبوبیت فیلم استفاده می‌کنیم. ما محبوبیت یک فیلم را به شش کلاس سوپر دوپر (SDH)، سوپر ضربه (SH)، ضربه (H)، بالاتر از میانگین (AA)، متوسط (A) و فیوپ (F) تقسیم کرده‌ایم. در مرحله بعد، در مرحله سوم و آخر، ماژولی برای تخمین مخاطب هدف ایجاد می کنیم. گروه مخاطب را به چهار گروه سنی خردسالان، نوجوانان، میانسالان و بزرگسالان تقسیم کرده ایم.

ما از یک مجموعه فیلم مشابه از ماژول اول استفاده کردیم و یک مجموعه ویژگی جدید از هر گروه سنی با در نظر گرفتن رتبه بندی فیلم و اطلاعات رای ایجاد کردیم. با استفاده از مجموعه داده های جدید، مدلی را با استفاده از میانگین های فازی و شباهت کسینوس برای تخمین محبوبیت فیلم در بین تمام گروه های سنی ساختیم. ادبیات مقاله به شرح زیر است.

۱. این کار تحقیقاتی یکی از مهمترین کارها در مطالعات قبلی بوده که از یک سیستم توصیه برای پیش‌بینی محبوبیت فیلم در مراحل اولیه تولید استفاده می‌کند.

۲. مدلی برای تخمین محبوبیت یک فیلم آینده در میان گروه های سنی مختلف با استفاده از میانگین فازی و فاصله کسینوس پیشنهاد کرد.

بقیه این مقاله به شرح زیر تنظیم شده است. بخش دوم کارهای مربوط به پیش بینی RS و فیلم را خلاصه می کند. بخش سوم چارچوب پیشنهادی ما را به طور مفصل تشریح می‌کند، تمام ویژگی‌ها را نشان می‌دهد و معیار موفقیت فیلم را معرفی می‌کند. در بخش چهارم مدل پیشنهادی به طور مفصل توضیح داده شده است. بخش پنجم نتایج تجربی را همزمان با مطالعه تطبیقی سایر مدل‌های آماری نشان‌داده شده و توضیح داده، مشارکت‌های پژوهشی و محدودیت‌های آنها و جهت‌گیری‌های تحقیقاتی بیشتر در بخش ششم ارائه می‌کند.

۲- کارهای مرتبط

در این بخش به بررسی دقیق آثار مرتبط با گذشته پرداخته ایم. کار ما به سه بخش مرتبط تقسیم می شود. برخی از مدل‌های پیشنهادی قبلی را بررسی و سپس پیش‌بینی محبوبیت فیلم را مورد بحث قرار داده ایم. در کار پیشنهادی، توصیه فیلم و محبوبیت فیلم به هم مرتبط نیستند. کار پیشنهادی ما از فیلم‌های پیشنهادی برای پیش‌بینی محبوبیت فیلم در آینده و پیش‌بینی مخاطب هدف فیلم استفاده کرده است. سیستم توصیه اصولاً به سه بخش تقسیم می‌شود [۱] [۳]، [۱۸]، فیلتر مشارکتی ([[8] [4] (CF، 17]]، فیلتر مبتنی بر محتوا ([8] (CBF، [9]، [13] [11]، 31]] و فیلترینگ ترکیبی [۱۴] [۱۶]. CF رویه‌ای است که می‌تواند مواردی را که کاربر ممکن است بر اساس پاسخ‌های کاربران مشابه ترجیح دهد، بازیابی کند. گروه وسیعی از افراد را جستجو می کند و دایره کوچکتری از کاربران با سلیقه های قابل مقایسه با یک کاربر خاص به دست می آورد. به چیزهایی که دوست دارند نگاه می‌کند و آنها را به هم متصل می‌کند تا فهرستی رتبه‌بندی شده از پیشنهادات را تشکیل دهد.

۲-۱- توصیه مبتنی بر محتوا

در برخی موقعیت‌های کاربردی، آیتم‌های توصیه‌شده باید از نظر محتوا با یک آیتم مرجع قابل مقایسه باشند، به عنوان مثال، برای توصیه‌های مورد مشابه [۱۹]. همچنین، اطلاعات محتوا توصیف بهتر [۲۰] را امکان پذیر می کند، که اغلب در سیستم های توصیه گر منصفانه حیاتی است. سیستم‌های توصیه‌گر مبتنی بر محتوا از اطلاعات فراداده اقلام یا موارد متنی استفاده می‌کنند [۲۱]. شروع داده های باز پیوندی (LOD) ایده های جدیدی را برای گسترش اطلاعات مورد با منابع دانش خارجی پیشنهاد می کند [۲۲]، [۲۳].

توصیه فیلم با استفاده از CBF یکی از پارادایم های تحقیقاتی پرکاربرد است. یک توصیه‌کننده فیلم مبتنی بر محتوا پیشنهاد شده‌است که در آن کاربران از ویژگی‌های فیلم استفاده می نمایند [۲۴]. رتبه بندی فیلم پیشنهادی با استفاده از مجموعه ویژگی فیلم انجام می شود. سیستم توصیه‌شده فیلم مبتنی بر محتوا، ویژگی‌های مختلف فیلم مانند ژانر فیلم، نام بازیگران، نام کارگردانان و سایر ویژگی‌ها را برای ساختن یک سیستم توصیه‌گر در نظر می‌گیرد. ژانر فیلمی که کاربران ترجیح می دهند تماشا کنند برای ساختن یک سیستم توصیه کننده با استفاده از مجموعه داده های فیلم [۲۵] استفاده شده است. همبستگی بین محتوا یا ویژگی ها برای از بین بردن شباهت بین موارد اندازه گیری می شود. یک شبکه چند ویژگی برای محاسبه همبستگی ها برای توصیه آیتم ها به کاربران پیشنهاد شده است [۲۶]. شباهت بین آیتم های همبسته مستقیم یا غیرمستقیم با استفاده از تحلیل شبکه محاسبه می شود. آنها یک مدل ترکیبی پیشنهاد کرده‌اند که در آن از برچسب‌های ژنومی فیلم با CBF برای توصیه فیلم‌هایی با سلیقه‌های مشابه استفاده شده است [۲۷]. مدل پیشنهادی پیچیدگی محاسباتی را با استفاده از تحلیل مؤلفه اصلی (PCA) و روش‌های همبستگی پیرسون برای کاهش برچسب‌های

اضافی و توزیع واریانس کم کاهش می‌دهد [۶۴]. در کار زیر، نویسندگان از شکاف بین ویژگی‌های سطح بالا و سطح پایین استفاده کرده‌اند [۲۸]. آن‌ها از رنگ‌های ویژگی سطح پایین، حرکت، بیش از حد و نور فیلم برای ایجاد یک سیستم توصیه ترکیبی استفاده کرده‌اند. یک سیستم توصیه فیلم جدید پیشنهاد شده و به مشکل شروع سرد برای آیتم جدید می پردازد [۲۹]. آنها توضیحات صوتی و تصویری استخراج شده از فیلم ها را ارائه کرده اند و ژنوم ویدئویی را توسعه داده اند. یک سیستم توصیه فیلم ترکیبی برای ترکیب تحلیل احساسات با ثبت مشارکتی (CF) پیشنهاد شده است [۳۰]. توییت های فیلم از سایت های وبلاگ نویسی برای درک احساسات عمومی، روندهای فعلی و پاسخ کاربران استفاده شده است. پراکندگی داده ها یکی از چالش های مهم برای الگوریتم های سیستم توصیه است. در کار زیر، یک شبکه متخاصم مولد (GAN) به بررسی پراکندگی داده‌های و رتبه‌بندی پرداخته است. آنها با بررسی رتبه بندی و شبکه های متخاصم مولد (RRGAN) پیشنهاد کرده اند که یک چارچوب ابتکاری برای این توصیه است [۶۷]. همچنین از GAN برای رتبه بندی فیلم بر اساس ترجیح کاربران استفاده شده است. در کار زیر، Lambda GAN برای توصیه فیلم های برتر استفاده شده است [۶۸].

۲-۲- پیش بینی موفقیت فیلم

پخش شاهکارهای فیلم یک مشکل شناخته شده تحقیقاتی است. این مشکل به طور کلی بر اساس زمان پیش بینی به دو گروه اصلی تقسیم می شود. کار مهمی پیشنهاد شده است که پیش‌بینی‌ها بسیار دیر در مرحله تولید قبل از اکران فیلم یا درست بعد از اکران فیلم [۳۲] [۳۸]، [۶۰] انجام می‌شد. کارهای محدودی انجام شده است که پیش‌بینی موفقیت فیلم در مرحله اولیه یا مرحله اولیه تولید اجرا شده است [۳۹] [۴۳]. پیش‌بینی دیرهنگام ممکن است با ویژگی‌های بیشتر فیلم برای افزایش دقت پیش‌بینی تسهیل شود. از سوی دیگر، برای پیش‌بینی اولیه، تنها چند ویژگی برای پیش‌بینی فیلم موجود است که مشکل را بسیار دشوارتر می‌کند.

یکی از مهم‌ترین بخش‌های مشکل ما، تعیین موفقیت یک فیلم است. هیچ مدل معیاری وجود ندارد که موفقیت یک فیلم را نشان دهد. تعداد کمی از آثار بر کل جعبه درآمد CE تمرکز کرده اند [۴۰]، [۴۳] [۴۸]. در همان زمان، برخی تعداد پذیرش را پذیرفته اند [۳۴]، [۴۹]. مفروضات اساسی برای کسب درآمد یا تعداد پذیرش به عنوان پارامتر موفقیت در نظر گرفته شده است. برخی از کارهای قبلی موفقیت را به عنوان توانایی حرفه ای ارزیابی می کردند. ممکن است مقدار عددی درآمد [۵۰] یا بازگشت سرمایه ([39] (ROI،]51[52]] باشد.

چندین اثر، فیلم‌ها را به دو دسته (موفق یا عدم موفقیت) توزیع کردند و دسته‌بندی‌های دودویی را انتخاب کردند. برخی پیش‌بینی را یک مشکل طبقه‌بندی چند کلاسه در نظر گرفتند و سعی کردند فیلم ها را به چند کلاس گسسته طبقه‌بندی کنند [۴۷]. پیش‌بینی‌هایی نیز بر روی مقادیر انتگرال پیوسته متریک‌های حرفه‌ای [۳۲]، [۳۹]، [۵۳] انجام می‌شود که مقادیر این معیارها در برخی آثار حاوی لگاریتمی است [۴۸]، [۵۰]، [۵۴].

مدل‌های یادگیری ماشینی قابل اعتماد پیش‌بینی فیلم با در نظر گرفتن این تکنیک‌های یادگیری، مدل‌های پیش‌بینی را با سطوح دقت معقولی توسعه داده‌اند [۵۵، [۵۲]، [۵۶]، [۶۵]. به عنوان مثال، [۵۶] برخی از مدل های یادگیری ماشین مانند تجزیه و تحلیل متمایز، رگرسیون لجستیک (LR)، درخت تصمیم (DT) و شبکه عصبی (ANN) را ارائه کرده و عملکرد را برای پیش بینی موفقیت یک فیلم اندازه گیری کرده است. نویسندگان [۵۷] معماری انتشار چند لایه پشتی و مدل شبکه عصبی توصیه با کیفیت بیشتری را پیشنهاد کرده اند که توسط [۵۶] ارائه شده است. نویسندگان IMDb [۵۸] و استراتژی‌های یادگیری ماشینی مانند ماشین بردار پشتیبانی (SVM) و رگرسیون خطی و لجستیک به‌دست آوردند. نویسندگان[۳۸] مدل گروه سینما (CEM) را برای افزایش دقت پیش‌بینی، شامل هفت مدل یادگیری ماشین و تمرکز بر انتخاب ویژگی‌ها معرفی کردند. تحقیق [۴۰] چند ویژگی جدید را برای پیش‌بینی موفقیت یک فیلم پیشنهاد کرد. آنها یک سیستم رأی گیری را برای پیش بینی با میانگین گیری خروجی از کلاس های مختلف یادگیری ماشین اتخاذ کرده اند.

در بیشتر آثار، فیلم‌های مشابه با استفاده از RS از فیلم موجود محاسبه می‌شوند. سیستم توصیه برای پیش بینی فیلم محبوب در آینده، پراکنده است. فیلم RS مبتنی بر محتوا می تواند برای پیدا کردن فیلم های مشابه برای فیلم آینده استفاده شود. اطلاعات جعبه‌ای از تمام این فیلم‌های مشابه می‌تواند برای تحلیل فیلم آینده استفاده شود. کار تحقیقاتی ما از فیلم RS مبتنی بر محتوا برای یک فیلم آینده و فیلم‌های مشابه استفاده کرده است. ما خروجی RS را تجزیه و تحلیل می کنیم و با موفقیت مدلی برای پیش بینی محبوبیت فیلم می سازیم. باز هم یک قدم جلوتر می رویم و همچنین مخاطبان هدف را از روی اطلاعات RS پیش بینی می کنیم. بیشتر کارهای تحقیقاتی

بر روی مسئله پیش‌بینی ضربه فیلم به عنوان یک مسئله طبقه‌بندی باینری متمرکز بودند. تعداد بسیار کمی از آثار [۴۰] مشکل را به عنوان یک مشکل چند طبقه ای حل می کنند، اما آنها دقت را در این فرآیند قربانی کرده اند. کار تحقیقاتی ما مسئله محبوبیت فیلم را در شش کلاس مختلف طبقه بندی کرده و به دقت بالایی دست یافته است. حداکثر آثار فقط مشکل پیش‌بینی محبوبیت فیلم را هدف قرار می‌دهند. آنها فقط محبوبیت فیلم را پیش بینی کرده اند.

۳- مواد و روش ها

این پژوهش با هدف توسعه، مدلی پیشنهاد شده که محبوبیت فیلم و اولویت سنی را با استفاده از توصیه‌های فیلم پیش‌بینی می کند. هدف ما طبقه بندی محبوبیت فیلم در بین شش کلاس {SDH، SH، H، AA، A، F} در مراحل اولیه تولید فیلم است. در مرحله بعد، هدف ما شناسایی مخاطبان هدف فیلم و تعیین تأثیر آن بر گروه های مخاطب است. مخاطبان را از نظر جمعیت‌شناسی در چهار گروه سنی (جوانان، نوجوانان، میانسالی، سالمندان) گروه‌بندی کرده که خروجی نهایی ما از این سیستم، پیش‌بینی محبوبیت فیلم از نظر سنی خواهد بود.

در این مطالعه، ما از یک سیستم توصیه فیلم مبتنی بر محتوا برای یافتن فیلم مشابه استفاده کردیم. در مرحله بعد از اطلاعات رای گیری و رتبه بندی هر فیلم پیشنهادی استفاده می کنیم. همه این داده ها برای آموزش مدل یادگیری عمیق CNN یک بعدی استفاده می شود. خروجی مدل CNN، طبقه بندی lm در بین شش کلاس است. ما میزان محبوبیت فیلم را پیش بینی می کنیم. در این مطالعه، ما از یک سیستم توصیه فیلم مبتنی بر محتوا برای یافتن فیلم مشابه استفاده کردیم. در مرحله بعد از اطلاعات رای گیری و رتبه بندی هر فیلم پیشنهادی استفاده می کنیم. همه این داده ها برای آموزش مدل یادگیری عمیق CNN یک بعدی استفاده می شود. خروجی مدل CNN، طبقه بندی lm در بین شش کلاس است. ما میزان محبوبیت فیلم را پیش بینی می کنیم.

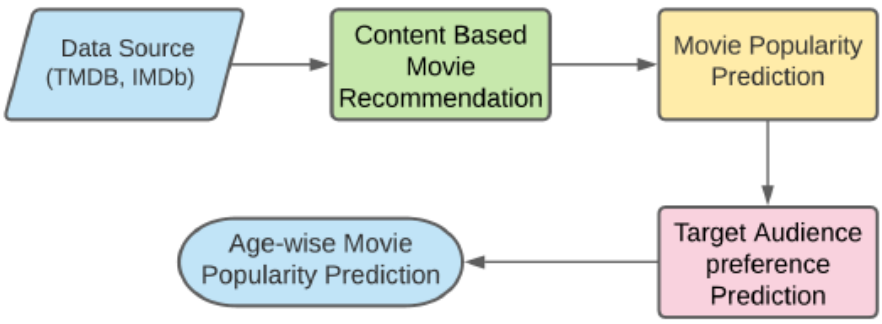
چارچوب کار ما دارای سه مرحله مهم است که در شکل ۱ در زیر آمده است.

۱) داده های فیلم و ویژگی های ذاتی فیلم را از مجموعه داده های TMDb بدست آورده و فیلم مشابه را با استفاده از یک سیستم توصیه فیلم مبتنی بر محتوا محاسبه کرده ایم.

۲) از اطلاعات مشابه فیلم و داده های رأی گیری از مجموعه بتای IMDb استفاده کرده و محبوبیت فیلم را با استفاده از رویکرد یادگیری عمیق پیش بینی نموده ایم.

۳) پیش بینی مخاطب هدف را با استفاده از ابزارهای فازی محاسبه نموده ایم.

در ادامه ما مجموعه داده جدیدی را معرفی کرده‌ایم که حاوی اطلاعات رأی‌گیری و رتبه‌بندی فیلم‌های پیشنهادی است که برای پیش‌بینی کلاس محبوبیت فیلم استفاده می‌شود. ما برای هر گروه سنی پارامترهای جدیدی به نام مرکز جهانی ارائه کرده ایم. ما همچنین رویکرد جدیدی را برای تخمین میزان علاقه یا محبوبیت یک فیلم در میان گروه‌های سنی مختلف ارائه کرده‌ایم.



شکل ۱: چارچوب گردش کاری.

الف) توصیف مجموعه داده

سیستم پیشنهادی دارای سه ماژول است. اولین ماژول یک سیستم توصیه مبتنی بر محتوا (CB) است (RS)، اولین ماژول که از پایگاه داده TMDb استفاده می کند. ماژول های دوم و سوم از پایگاه داده IMDb استفاده می کنند. چندین پایگاه داده عمومی در بازار موجود است و همه این پایگاه های داده برای توصیه فیلم و سیستم پیش بینی محبوبیت فیلم استفاده می شوند.

سیستم پیشنهادی فیلم مبتنی بر محتوا از مجموعه داده‌های tmdb_5000_movies و tmdb_5000_credits استفاده می‌کند که برای عموم در دسترس هستند [۶۲]. ماژول پیش‌بینی ضربه فیلم و پیش‌بینی مخاطب هدف از مجموعه داده رتبه‌بندی IMDb استفاده می‌کند. استفاده از دو پایگاه داده مختلف (TMDb و IMDb) مشکلات همگام سازی ایجاد می کند زیرا آنها از دو ID فیلم جداگانه استفاده می کنند. سیستم پیشنهادی از مجموعه داده کوچک پیوند برای ادغام دو پایگاه داده استفاده می کند.

مجموعه داده tmdb_5000_movie با داده های فیلم ۴۸۰۳ سازگار است. این شامل ۲۰ ویژگی فیلم است. که حسابداران آن فیلم با سال انتشار از ۱۹۱۶ تا ۲۰۱۷ است. از بین ۲۰ ویژگی، ما فقط ۴ ویژگی را انتخاب کرده ایم. ویژگی های انتخاب شده عبارتند از کلمات کلیدی، نمای کلی، شعار و ژانر. ویژگی هایی مانند بودجه، درآمد و سال اکران بسیار به زمان بستگی دارد، و از آنجایی که ما فیلمی با بازه بیش از صد سال را در نظر می گیریم، این ویژگی ها قابل انتخاب نیستند. سایر ویژگی‌ها مانند زمان اجرا، زبان گفتاری، عنوان اصلی، صفحه اصلی نیز به سیستم پیشنهادی بی‌ربط هستند.

مجموعه داده tmdb_5000_credits از ۴۸۱۳ داده فیلم تشکیل شده است. مجموعه داده دارای ۴ ویژگی movie_id، عنوان، بازیگران و خدمه است. ما نام سه بازیگر اصلی و نام کارگردان هر فیلم را از مجموعه داده ها استخراج کرده ایم. کارگردان یکی از بی نظیرترین شخصیت های یک فیلم است. به طور مشابه، ۳ بازیگر اول نیز بسیار مهم هستند. ویژگی های فیلم انتخاب شده برای سیستم فیلم مبتنی بر محتوا پیشنهادی در جدول ۱ نشان داده شده است.

جدول ۱. نمونه ای از داده های فیلم مورد استفاده در سیستم توصیه CB.

Attribute	Value
Movie_id	68721
Genres	Action, Adventure, Science Fiction
Keywords	Terrorist, war on terror, Tennessee, superhero, tony stark, war machine, extremis etc.
Overview	When Tony Stark's world is torn apart by a formidable terrorist called the Mandarin, he starts an odyssey of rebuilding and retribution.
Tagline	Unleash the power behind the armor.
Casts	Robert Downey Jr, Gwyneth Paltrow, Don Cheadle
Director	Shane Black

ب) پیش پردازش مجموعه داده

ویژگی های مورد استفاده در سیستم پیشنهادی فیلم CB در جدول ۱ ذکر شده است. ارزش ویژگی هایی مانند ژانر، بازیگران، کارگردان و کلمات کلیدی در داده های JSON وجود دارد. داده‌های JSON را به رشته تبدیل کرده و سپس تمام ابرداده‌ها را حذف و فقط داده های حاوی مقادیر باقیمانده است. در مرحله بعد، برای هر ویژگی، یک لیست با تمام مقادیر منحصر به فرد ویژگی های ذکر شده در مجموعه داده ایجاد شده است. در مرحله بعد، رمزگذاری یکباره را برای همه نگرش ها برای هر فیلم انجام می دهد. شعار یک فیلم با ویژگی نمای کلی ضمیمه شده است. ترکیب این دو ویژگی باید چندین مرحله را طی کند تا شباهت بین آنها مشخص شود.

مرحله ۱: داده ها را به درستی پاک کنید. علاوه بر این، این به ما امکان می دهد جملات، پاراگراف ها و در نهایت اسناد را به مجموعه ای از کلمات واحد کاهش دهیم.

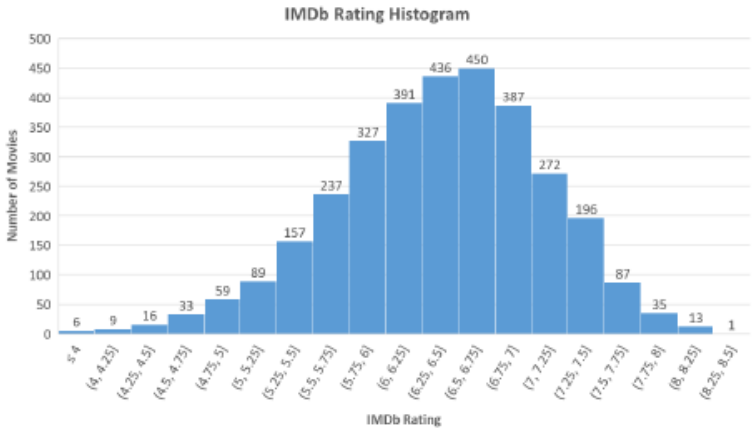
مرحله ۲: از طریق فرآیند ریشه‌یابی غلبه بر واژه‌های برآمده به شکل ریشه، ریشه یا مبدأ کلمه آنها بروید.

مرحله ۳: کلمات توقف را از مجموعه کلمات حذف کنید. فرهنگ لغت انگلیسی برای حذف کلمات توقف استفاده شده است.

مرحله ۴: پس از ساخت مجموعه داده های پاک و فیلتر، بعدی اجرای عملکرد اصلی است. فراوانی ترم با تعداد دفعاتی که عبارت t در سند d تکرار می شود محاسبه می شود.

ج) برچسب‌گذاری محبوبیت فیلم (HIT TO FLOP)

تصویر متحرک یکی از شاخه های هنر است. چندین پارامتر با صنعت فیلم مرتبط است. علاقه به فیلم یک صنعت پیچیده و گسترده است که بسیاری از عناصر با این صنعت مرتبط هستند. دیدگاه‌های متفاوتی وجود دارد تا پارامترهای ارزیابی موفقیت یک فیلم را در نظر بگیریم. درآمد باکس آفیس می تواند یکی از این پارامترها باشد. بودجه و درآمد فیلم متغیر بوده و به صنعت سینما بستگی دارد. برای برخی از صنایع، بودجه ممکن است به طور کلی بیشتر از سایر صنایع کوچک فیلم باشد یا از فیلمی به فیلم دیگر متفاوت باشد. تعریف محدوده درآمد [۴۰] برای طبقه بندی موفقیت فیلم در مورد همه فیلم ها صدق نمی کند. همچنین تثبیت حاشیه سود [۳۹] مشکل طبقه بندی را حل نخواهد کرد. مقیاس حاشیه سود بدون شک برای فیلم های کم هزینه کمتر و برای فیلم های با بودجه بالا بالاتر است. با در نظر گرفتن همه این موارد، رتبه IMDb یکی از معیارهای مهم برای پیش بینی موفقیت فیلم است و همچنین رتبه IMDb یک رتبه پذیرفته شده جهانی است. شکل ۲ هیستوگرام رتبه بندی IMDb همه فیلم های در نظر گرفته شده در پایگاه داده ما را نشان می دهد. این نشان می دهد که رتبه بندی نزدیک به یک توزیع نرمال است و به استحکام پیش بینی مدل کمک می کند. در کار خود، از رتبه بندی IMDb به عنوان پارامتر اصلی برای تعیین موفقیت فیلم استفاده کرده ایم.



شکل ۲. توزیع شماره فیلم را در رتبه بندی IMDb نشان می دهد.

در این چرخه کاری محبوبیت فیلم در شش طبقه، طبقه بندی شده است. در مسئله طبقه‌بندی فیلم در مراحل اولیه تولید، ماژول پیش‌بینی رتبه‌بندی IMDb را برای طیفی از رتبه‌بندی‌های IMDb پیش‌بینی می‌کند. با توجه به رتبه بندی پیش بینی شده در IMDb، فیلم آینده e در شش کلاس طبقه بندی می شود.

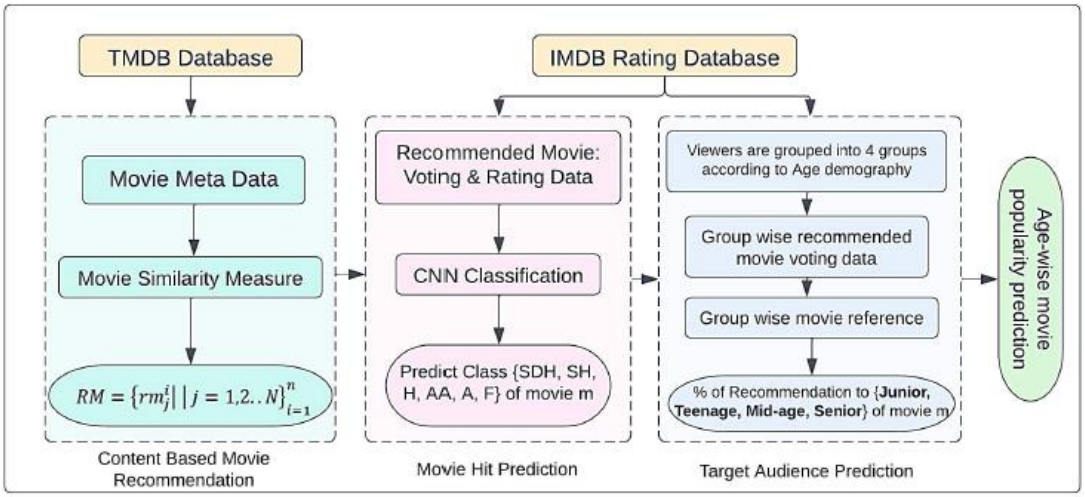
کلاس ها به عنوان ضربه فوق العاده دوپر (SDH)، ضربه فوق العاده (SH)، ضربه (H)، بالاتر از میانگین (AA)، متوسط (A) و فلاپ (F) مشخص می شوند. مجموعه داده های فیلم را آماده کرده و طبق رتبه بندی IMDb فیلم را در کلاس های مختلف برچسب گذاری کرده ایم. جدول ۲ نحوه طبقه بندی فیلم ها را بر اساس رتبه بندی IMDb به کلاس های مختلف نشان می دهد.

جدول ۲. طبقه بندی فیلم بر اساس رتبه بندی IMDb.

IMDb Rating	0 – 4.9	5 – 5.9	6 – 6.9	7 – 7.9	8 – 8.9	9 - 10
Movie Class	F	A	AA	H	SH	SDH

۴- کار پیشنهادی

سیستم پیشنهادی دارای سه ماژول اصلی مرتبط با یکدیگر است. ماژول اول یک مدل سیستم توصیه فیلم مبتنی بر محتوا است که ۱۰ فیلم مشابه از همه فیلم های فهرست شده در مجموعه داده را تولید می کند. ماژول دوم یک ماژول پیش بینی ضربه فیلم است. خروجی ماژول اول ورودی ماژول دوم است. در مرحله بعد، ماژول سوم مخاطبان را بر اساس سن آنها گروه بندی می کند و مناسب ترین گروه مخاطب هدف را برای هر فیلم پیش بینی می کند. در نهایت، کل سیستم یک پیش‌بینی محبوبیت فیلم از نظر سنی از فیلم آینده ارائه می‌کند. نمودار کلی جریان فرآیند در شکل ۳ نشان داده شده است.



شکل ۳. نمودار کلی جریان فرآیند.

الف) توصیه فیلم مبتنی بر محتوا

از فیلتر مبتنی بر محتوا برای یافتن یک فیلم مشابه استفاده می شود. که از ویژگی های فیلم برای یافتن شباهت بین این دو فیلم استفاده می کند. اجازه دهید ویژگی $F = \{F_1, F_2, F_3, \ldots, F_m\}$ تنظیم کند. شباهت بین هر دو فیلم m_i & m_j را در مورد ویژگی F_k به شرح ذیل محاسبه می شود:

$$dist_F_{k_{i,j}} = similarity(F_{k_i}, F_{k_j}) \qquad (1)$$

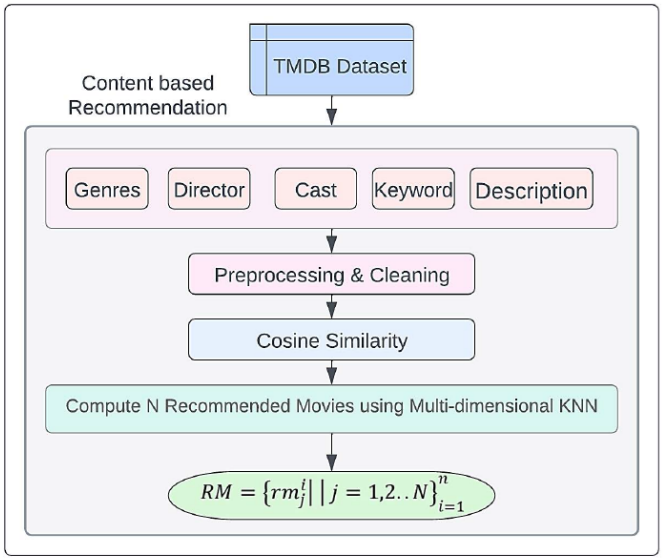
$$dist_{i,j} = \cup_{k=1}^{m} \{dist_{F\,k_{i,j}}\} \qquad (2)$$

در (۱)، $dist_{ij}$ بردار فاصله بین دو فیلم m_i & m_j است. هدف این ماژول محاسبه N بیشتر فیلم های مشابه فیلم m_i است. شباهت بین هر دو فیلم با استفاده از ویژگی های مختلف اندازه گیری شده است. فاصله بین هر دو فیلم بردار بعد m است. الگوریتم نزدیکترین همسایه برای یافتن N بیشتر فیلم های مشابه استفاده شده است. اندازه گیری شباهت کلی بین دو فیلم با استفاده از معیارهای شباهت مانند شباهت کسینوس به شرح ذیل محاسبه می شود.

$$c_dist_{ij} = cosine_sim(dist_{ii}, dist_{ij}), \qquad (3)$$

$$c_dist_i = \{c_dist_{ij}\}_{j=1}^{n} \qquad (4)$$

در (۴)، ما اندازه گیری شباهت بین هر فیلم m_i را با تمام n فیلم دیگر موجود در مجموعه داده محاسبه کرده ایم. بردار فاصله بعدی m با استفاده از اندازه گیری شباهت کسینوس به فاصله یک بُعدی کاهش می یابد. شکل ۴ بلوک دیاگرام ماژول توصیه فیلم CB را نشان می دهد.



شکل ۴. ماژول توصیه فیلم.

ب) پیش بینی ضربه فیلم

هدف ماژول دوم در این مسئله تحقیقاتی ساختن یک سیستم پیش‌بینی ضربه فیلم است. این ماژول خروجی ماژول قبلی را می پذیرد. از مجموعه ماژول قبلی از فیلم توصیه شده هر فیلم، m_i به عنوان ورودی استفاده می شود، یعنی $RM = u_{i=1}^{n} \{rm_j^i | j = 1,2, \dots, N\}$ با آن مجموعه داده رتبه بندی IMDB به عنوان ورودی استفاده شده است. پیش‌بینی ضربه فیلم یک مشکل طبقه‌بندی چند طبقه است. داده های ورودی پردازش می شوند و در یک مدل یادگیری عمیق قرار می گیرند، و خروجی یک طبقه بندی چند کلاسه است. در این مسئله پژوهشی، فیلم‌ها به شش کلاس ضربه فوق العاده دوپر (SDH)، ضربه فوق العاده (SH)، ضربه (H)، بالاتر از میانگین (AA)، متوسط (A) و فلاپ (F) طبقه بندی می شوند. شکل ۴ چارچوب ماژول پیش بینی ضربه فیلم را نشان می دهد.

در این ماژول از امتیاز و جزئیات رای گیری N فیلم توصیه شده rm_j^i فیلم m_i استفاده کرده ایم. اجازه دهید $v_{r,k}^i$ تعداد رای را با درجه بندی r برای *kth* فیلم توصیه شده m_i نشان دهد. v_j^i نشان دهنده تعداد کل آرا با رتبه r برای همه فیلم های توصیه شده m_i است.

Algorithm 1 Movie recommendation using multi-dimensional KNN.

Input:
Movie Feature set F = $\{F_1, F_2, \ldots F_N\} = \{D, K, G, DT, C\}$
Movie Description & Tag line $D = \{d_i\}_{i=1}^{n}$
Movie Keywords $K = \{k_i\}_{i=1}^{n}$,
Movie Genres $G = \{g_i\}_{i=1}^{n}$
Movie Director $DT = \{dt_i\}_{i=1}^{n}$,
Movie Cast $C = \{c_i\}_{i=1}^{n}$
Output:
N most similar movie of movie $m_i, \forall m_i \{rm_j^i | j = 1, 2, \ldots, N\}_{i=1}^{n}$

1. $D_{TF-IDF} = \{d_ti\}_{i=1}^{n}$ calculate $TF - IDF \forall t_{ji} \in d_i$, where $d_i \in D$
2. $K_{BIN} = \{w_{bin}\}_{i=1}^{n}$
   where k_i transfer to binary word binw_bin_i
3. $G_{BIN} = \{g_{bin}\}_{i=1}^{n}$
   where g_i transfer to binary genrebing_bin_i
4. $DT_{BIN} = \{dt_{bin}\}_{i=1}^{n}$
   where dt_i transfer to binary director bindt_bin_i
5. $C_{BIN} = \{c_{bin}\}_{i=1}^{n}$
   where c_i transfer to binary cast binc_bin_i
6. **for** (*Each movie* $m_i \in M$)
7. **for** (*Each movie* $m_j \in M \& m_j \neq m_i$)
8. $dist_d_{ij} = cosine_sim\,(d_ti, d_tj)$
9. $dist_w_{ij} = cosine_sim\,(w_bin_i, w_bin_j)$
10. $dist_g_{ij} = cosine_sim\,(g_bin_i, g_bin_j)$
11. $dist_dt_{ij} = cosine_sim\,(dt_bin_i, dt_bin_j)$
12. $dist_c_{ij} = cosine_sim\,(c_bin_i, c_bin_j)$
13. $dist_{ij} = \{dist_d_{ij}, dist_w_{ij}, dist_g_{ij}, dist_dt_{ij}, dist_c_{ij}$
14. $dist_i = \{dist_{ij}\}_{j=1}^{n}$
15. **for**(*Each movie* $m_j \in M \& m_j \neq m_i$)
16. $c_dist_{ij} = cosine_sim(dist_{ii}, dist_{ij})$
17. $c_dist_i = \{c_dist_{ij}\}_{j=1}^{n}$
18. $c_dist_i = sort(\{c_dist_{ij}\}_{j=1}^{n})$
19. $RM^i = \{rm_j^i | j = 1, 2, \ldots, N = firstNmoviesof\, c_dist_i$
20. $RM = \{rm_j^i | j = 1, 2, \ldots, N\}_{i=1}^{n}$
21. **Return** RM

الگوریتم ۱. توصیه فیلم با استفاده از KNN چند بعدی.

$$V_r^i = \sum_{k=1}^{N} v_{r,k}^i \qquad (5)$$

$$V^i = \{V_r^i | r = 1, 2, \ldots, 10\} \qquad (6)$$

در (۶)، $\mathbf{v^i}$ جزئیات رای گیری فیلم m_i را نشان می دهد. در (۷) رتبه $\mathbf{r_j^i}$ از هر فیلم توصیه شده، نیز به صورت زیر در نظر گرفته شده است:

$$R^i = \{R_j^i \,|\, j = 1, 2, \ldots, N\}$$
$$V = \{V^i\}_{i=1}^{n} \quad and\ R = \{R^i\}_{i=1}^{n}$$
$$\text{creates input dataset} \qquad X = V \cup R \qquad (7)$$

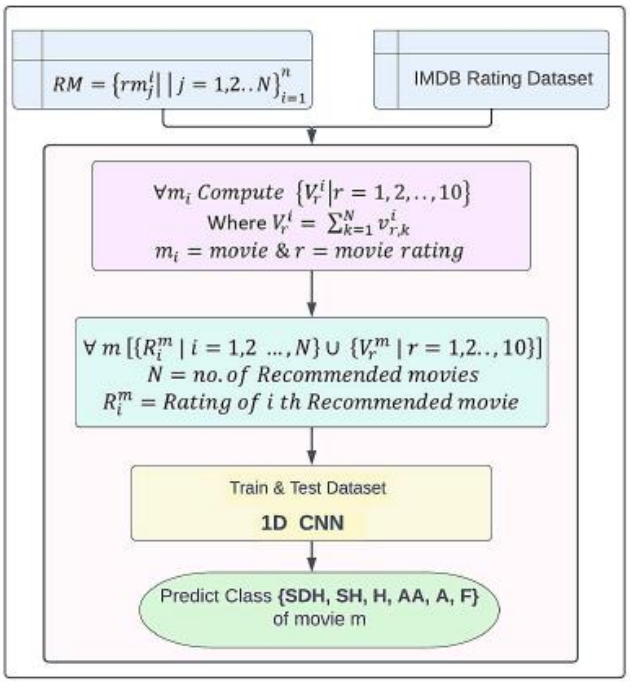
جدول ۳. نمونه ای از جزئیات رای گیری.

$\boldsymbol{V^i} = \{V_r^i \mid r = 1, 2, .., 10\}$	$m_i =$ Avatar	$m_i =$ Spider-Man 3
V_1^i	822428	210605
V_2^i	800110	206442
V_3^i	806286	482927
V_4^i	446746	561242
V_5^i	190478	324569
V_6^i	101340	154969
V_7^i	51821	67734
V_8^i	29602	34332
V_9^i	19301	19318
V_{10}^i	38270	26599

جدول ۴. نمونه ای از جزئیات رتبه بندی.

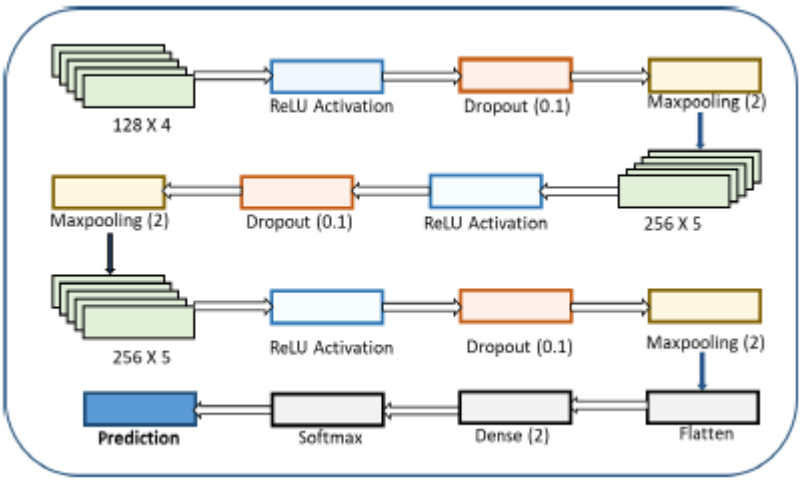
$\boldsymbol{R^i} = \{R_j^i \mid j = 1, 2, .., N\}$	$m_i =$ Avatar	$m_i =$ Spider-Man 3
R_1^i	7.1	6.7
R_2^i	7.3	6.8
R_3^i	7.7	5.7
R_4^i	7.7	6.3
R_5^i	6.8	7.6
R_6^i	6.3	5.3
R_7^i	7.5	5.8
R_8^i	5.4	5.8
R_9^i	6.5	6.9
R_{10}^i	7.5	4.8

جدول ۳ و ۴ مقادیر ویژگی های $\mathbf{v}^i$ و رتبه بندی $\mathbf{r}^i$ دو فیلم (آواتار و مرد عنکبوتی ۳) را به عنوان مثال ارائه می دهد. برای پیش بینی کلاس فیلم، از مجموعه داده X به عنوان ورودی استفاده شده است. مجموعه داده های ورودی شامل جزئیات داده ای ۳۱۰۰ فیلم می باشد. مجموعه داده با شش کلاس مختلف برچسب گذاری شده است. مجموعه داده های ورودی دوباره به دو بخش آموزش و تست تقسیم می شود. بخش آموزشی شامل ۲۳۲۵ داده فیلم و بخش تست نیز شامل ۷۷۵ داده فیلم است. مدل طبقه‌بندی شبکه عصبی پیچیدگی برای طبقه‌بندی فیلم‌ها به ۶ کلاس مختلف بر اساس محبوبیت استفاده شده است. شکل ۵ بلوک دیاگرام ماژول پیش بینی ضربه فیلم را نشان می دهد.



شکل ۵. مدل پیش بینی ضربه فیلم.

از ۱۲۸ فیلتر در لایه اول با اندازه هسته ۵ استفاده شده است. با آن، max-pooling برابر ۲ تخمین زده می‌شود. لایه بعدی ۱۲۸ اندازه هسته و همان تابع فعال‌سازی Relu و حذف ۰.۱ را حفظ می‌کند و این کار را دو بار تکرار می‌کند. در نهایت از یک لایه فلاترینگ استفاده شده است. Last Layer یک لایه متراکم با شش کلاس در اینجا با ۶ کلاس پیش بینی است. چندین بهینه ساز Keras مانند Adam، SGD، RMSprop آزمایش شده اند و در نهایت به دلیل دقت بهتر، بهینه ساز RMSprop را انتخاب کرده ایم. شکل ۶ توپولوژی 1D-CNN پیشنهادی را نشان می دهد.



شکل ۶. توپولوژی 1D-CNN پیشنهادی

ج) پیش بینی مخاطبان هدف

سومین و آخرین ماژول، ترجیحات مخاطب هدف را با توجه به سن جمعیت شناسی پیش بینی می کند. برای پیش‌بینی مخاطبان فیلم، ما همه فیلم‌های پیشنهادی ارائه‌شده توسط اولین ماژول خود را در نظر گرفته‌ایم.

Algorithm 2 Movie popularity prediction using deep learning.

Input:
Recommended Movie details $RM = \{rm_j^i | j = 1, 2, \ldots, N\}_{i=1}^{n}$
Movie Rating Database M_r
Output:
Prediction of Popularity Class (0, 1, 2, 3, 4, 5) for *movie* m_i

1. **for**(*Each movie* $m_i \in M$)
2. **for**(*Rating* $r = 1, 2, \ldots, 10$)
3. $V_r^i = \sum_{k=1}^{N} v_{r,k}^i$,
4. *where* $v_{r,k}^i$ = no. of vote with rating r of *kth* recommended movie of m_i
5. $V^i = \{V_r^i | r = 1, 2, \ldots, 10$ voting details of each movie m_i
6. $V = \{V^i\}_{i=1}^{n}$ voting details of all movie M
7. $R = \{R_j^i | j = 1, 2, \ldots, N\}_{i=1}^{n}$, where R_j^i = rating of *jth* recommended movie of m_i
8. *Create Dataset* $X = V \cup R$
9. $Y = \{y_i | y_i \in (0, 1, 2, 3, 4, 5)\}_{i=1}^{n}$
10. $X = X_{train} \cup X_{test}$
11. Y = Y_train ∪ Y_test
12. Y ^_test = CNN(X_train,Y_train,X_test)
13. **Return** Y ^_test

الگوریتم ۲. پیش بینی محبوبیت فیلم با استفاده از یادگیری عمیق.

این سیستم از رتبه‌بندی کاربران توسط هر گروه سنی استفاده می‌کند و تعداد جزئیات رأی از هر گروه را به تجزیه و تحلیل سیستم فیلم توصیه‌شده از تمام اطلاعات رأی‌گیری و رتبه‌بندی از هر گروه از همه فیلم‌های پیشنهادی می‌گیرد. در نهایت، این ماژول محبوبیت فیلم آینده را برای هر گروه سنی پیش بینی می کند. ماژول پیش بینی مخاطب هدف ورودی را از خروجی ماژول اول دریافت می کند و همچنین مجموعه داده های رتبه بندی IMDB را به عنوان ورودی می گیرد.

ماژول توصیه فیلم $(rm_j^i | j = 1,2, \ldots, N)$ فیلم‌های مشابه از یک فیلم مشخص را تولید می‌کند. ما از مجموعه ای از فیلم های توصیه شده از همه فیلم های $RM = u_{i=1}^{n} \{rm_j^i | j = 1,2, \ldots, N\}$ موجود در مجموعه داده استفاده نموده ایم. مجموعه داده‌های رتبه‌بندی IMDB، از جمله اطلاعات رای‌گیری و رتبه‌بندی همه فیلم‌های موجود در مجموعه داده را نیز گرفته‌ایم. سیستم پیشنهادی مخاطبان را بر اساس جمعیت سنی به چهار گروه تقسیم می کند. سیستم پیشنهادی پیش‌بینی می‌کند که فیلم برای هر گروه برای یک فیلم چقدر ارجحیت دارد. کدام گروه بیشتر فیلم را ترجیح می دهند و کدام گروه فیلم را دوست ندارند. جدول ۵ گروه های سنی بینندگان را نشان می دهد. ما هر گروه Gr_i را جدا کرده ایم. علاوه بر این، با استفاده از مجموعه داده فیلم توصیه شده و مجموعه داده رتبه بندی IMDB، مجموعه داده جداگانه ایجاد کرده که هر مجموعه داده شامل اطلاعات رتبه‌بندی و رأی‌گیری همه فیلم‌های توصیه‌شده یک فیلم m_i است.

Age	Group	Group name
0-17	Gr-1	Junior
18-29	Gr-2	Teenage
30-44	Gr-3	mid-age
45 +	Gr-4	Senior

جدول ۵. گروه سنی بینندگان

ما هر گروه Gr_j را جدا کرده ایم. علاوه بر این، با استفاده از مجموعه داده فیلم توصیه شده و مجموعه داده رتبه بندی IMDb، مجموعه داده جداگانه ایجاد کنید. هر مجموعه داده شامل اطلاعات رتبه‌بندی و رأی‌گیری همه فیلم‌های توصیه‌شده یک فیلم $\mathrm{m_i}$ است.

$$Gr_j^i = \{R^i_{Gr_j,k}, Vr^i_{Gr_j,k} | k = 1, \dots, N\} \qquad (8)$$

در (۸)، Gr_j^i نشان‌دهنده گروه j برای فیلم $\mathrm{m_i}$ است، جایی که $R^i_{Grj,k}$ میانگین رتبه‌بندی kth فیلم توصیه‌شده فیلم $\mathrm{m_i}$ از گروه j را توصیف می‌کند. $vr^i_{Grj,k}$ نشان دهنده نسبت رای دهی kth فیلم توصیه شده فیلم $\mathrm{m_i}$ از گروه j است. هر گروه Gr_j^i یک خوشه جداگانه ایجاد می کند. برای یک خوشه، Gr_j^i ترجیح گروه سنی J را برای یک فیلم معین توصیف می کند. برای هر خوشه، همه فیلم‌های توصیه‌شده $k = (1, \dots, N)$ و میانگین امتیاز و جزئیات رای آنها را برای گروه سنی Gr_j^i در نظر بگیرید. اگر امتیاز کلی و نسبت رای یک گروه $(Gr_j^i | \ j = 1,2,3,4)$ نسبتاً بالاتر از گروه دیگر باشد $(Gr_l^i | \ l = 1,2,3,4) \ l \neq j)$ آنگاه نشان دهنده فیلم $\mathrm{m_i}$ توسط گروه Gr_j^i و سپس Gr_l^i نسبتا ارجح است. جداول ۶ و ۷ رتبه بندی $R^i_{Grj,k}$ و $vr^i_{Grj,k}$ نظرسنجی دارای مقادیر گروه ۳ (میانسال) برای دو فیلم نمونه (آواتار و مرد عنکبوتی ۳) است.

جدول ۶. نمونه های از جزئیات رتبه بندی گروه ۳

$R^i_{Gr_3} = \{R^i_{Gr_3,k} \mid k = 1,2,..,N\}$	$m_i =$ Avatar	$m_i =$ Spider-Man 3
$R^i_{Gr_3,1}$	7.6	7.2
$R^i_{Gr_3,2}$	8.6	7.2
$R^i_{Gr_3,3}$	8.4	6.2
$R^i_{Gr_3,4}$	4.6	6.4
$R^i_{Gr_3,5}$	4.8	7.7
$R^i_{Gr_3,6}$	5.1	7.5
$R^i_{Gr_3,7}$	5.2	5.4
$R^i_{Gr_3,8}$	6	6.1
$R^i_{Gr_3,9}$	7.1	6
$R^i_{Gr_3,10}$	8.6	7.4

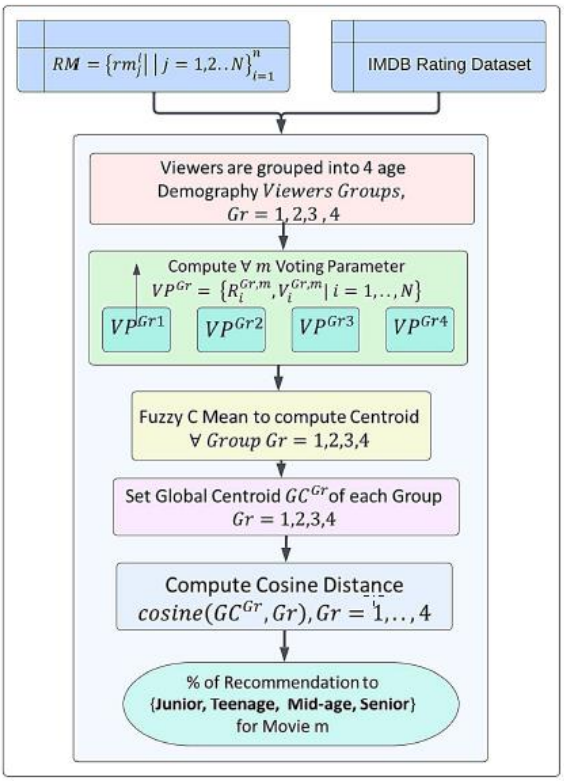
$$(C_R^i_{Gr_j}, C_V^i_{Gr_j})_{j=1}^4 = FCM(\{R^i_{Gr_j,k}, Vr^i_{Gr_j,k} | \times k = 1, \dots, N \ \& \ j = 1, \dots, 4\}) \qquad (9)$$

$$C_centroid^i_{Gr_j} = (C_R^i_{Gr_j}, C_Vr^i_{Gr_j}) \qquad (10)$$

برای مقایسه دو گروه، باید یک اولویت کلی برای هر گروه داشته باشیم. مرکز خوشه عملکرد یک خوشه را اندازه گیری می کند. هر گروه از چندین نقطه داده تشکیل شده است. برای تعیین مرکز خوشه ای هر گروه از میانگین c فازی استفاده می شود. مرکز خوشه پارامتری است برای اندازه گیری عملکرد یا علاقه گروه مخاطب به فیلم m. به طور تجربی ما یک مرکز جهانی برای همه فیلم ها تنظیم کرده ایم.

جدول ۶. نمونه های از جزئیات رتبه بندی گروه ۳

$V^i_{Gr_3} = \{V^i_{Gr_3,k} \mid k = 1,2,..,N\}$	$m_i =$ Avatar	$m_i =$ Spider-Man 3
$V^i_{Gr_3,1}$	76438	214612
$V^i_{Gr_3,2}$	431187	272023
$V^i_{Gr_3,3}$	275633	86073
$V^i_{Gr_3,4}$	9803	41716
$V^i_{Gr_3,5}$	19141	69057
$V^i_{Gr_3,6}$	24309	79096
$V^i_{Gr_3,7}$	19183	59705
$V^i_{Gr_3,8}$	38607	12752
$V^i_{Gr_3,9}$	55223	38607
$V^i_{Gr_3,10}$	458626	27175



شکل ۷. بلوک دیاگرام مدل پیش بینی مخاطب هدف

مرکز جهانی برای هر گروه منحصر به فرد است. فاصله مرکز خوشه از مرکز جهانی اندازه گیری می شود. شکل ۷ بلوک دیاگرام مدل پیش بینی مخاطب هدف را نشان می دهد.

$$G_centroid = (G_R, G_Vr) \qquad (11)$$

در بخش قبل، محبوبیت یک فیلم را بسته به رتبه بندی فیلم تعریف کردیم. اگر امتیاز فیلم $\mathrm{mr_i} > 7$ باشد، فیلم محبوب را تشخیص می دهد. در پارامتر سراسری مرکز خود، یک رتبه‌بندی جهانی $G_R = 7$ را نیز تعیین کرده‌ایم. ما تعداد رای‌گیری‌های انجام شده را بیشتر یا مساوی با مقدار متوسط در نظر گرفتیم. مقدار متوسط برای هر گروه متفاوت است. ما نسبت رای گیری را در نظر گرفته ایم، و نسبت رای گیری جهانی $G_Vr = 1$ را ثابت کرده ایم. سپس، با ادغام پارامترهای جهانی، قرار است شباهت بین مرکز جهانی و مرکز فیلم اندازه گیری کنیم.

$$similarity^{i}_{Gr_j} = cosine_sim(G_centroid, Movie_centroid^{i}_{Gr_j}) \qquad (12)$$

فاصله بین فیلم $\mathrm{C_centroid^{i}_{Grj}}$ از گروه جهانی $\mathrm{centroid}$، G_centroid محبوبیت فیلم را در گروه تعیین می کند. اگر فاصله کم باشد، پس محبوبیت زیاد است. اگر فاصله زیاد باشد، محبوبیت فیلم در گروه کم است. اندازه گیری فاصله به درصد محبوبیت تبدیل می شود. اگر یک مرکز گروه دارای $C_R^{i}_{Gr_i} \geq G_R$ and $C_Vr^{i}_{Gr_i} \geq G_Vr$ باشد، آنگاه معیار محبوبیت ۱۰۰٪ خواهد بود. در نهایت، این ماژول محبوبیت یک فیلم را در بین هر گروه سنی پیش‌بینی می‌کند.

Algorithm 3 Target audience preference prediction using fuzzy c mean.

Input:

Recommended Movie details $RM = \{rm_j^i | j = 1, 2, \ldots, N\}_{i=1}^{n}$

Movie Rating Database M_r

Output:

Preference percentage to each group $\{Gr_j | j = 1, 2, 3, 4\}$

1. Audience $Au = Gr_1 \cup Gr_2 \cup Gr_3 \cup Gr_4$
2. **for** (*Each movie* $m_i \in M$)
3. **for**(*Each group* $\{Gr_j | j = 1, 2, 3, 4\}$)
4. $VP_{Gr_j}^i = \{R_{Gr_j,k}^i\}_{k=1}^{N} \cup \{V_{Gr_j,k}^i\}_{k=1}^{N}$
5. $\{VP_{Gr_j} | j = 1, 2, 3, 4\}$ voting parameters for each group Gr_i
6. $VP_{Gr_j}^i$ for each movie m_i formed a cluster $Gr_j^i | j = 1, 2, 3, 4$
7. **for** (*Each movie* $m_i \in M$)
8. **for**(*Each group*$\{Gr_j | j = 1, 2, 3, 4\}$)
9. Elements in cluster $Gr_j^i = \left\{ R_{Gr_j,k}^i, V_{Gr_j,k}^i \middle/ median(\sum_{k=1}^{N} V_{Gr_j,k}^i) \,|\, k=1, \ldots, N \right\}$
10. $Gr_j^i = \{R_{Gr_j,k}^i, Vr_{Gr_j,k}^i | k = 1, \ldots, N$
11. $(C_R_{Gr_j}^i, C_V_{Gr_j}^i)_{j=1}^{4} = FCM(\{R_{Gr_j,k}^i, Vr_{Gr_j,k}^i | k = 1, \ldots, N \& j = 1, \ldots, 4\})$
Centroid of each cluster using Fuzzy c mean
12. $Movie_centroid_{Gr_j}^i = (C_R_{Gr_j}^i, C_Vr_{Gr_j}^i)$
13. Set global centroid $= G_centroid = (G_R, G_Vr)$
14. $similarity_{Gr_j}^i = cosine_sim(G_centroid, Movie_centroid_{Gr_j}^i)$
15. Calculate percentage similarity $Per_similarity_{Gr_j}^i$ of each group Gr_j
16. **Return** *Per_similarity*

الگوریتم ۳. پیش بینی اولویت مخاطب هدف با استفاده از معنی c فازی

جدول ۸. فیلم های پیشنهادی فیلم The Terminator.

Recommended Movie	Similarity distance	Genres
Terminator 2: Judgment Day	1.2698	Action, ScienceFiction, Thriller
True Lies	1.3677	Action, Thriller
The Abyss	1.3725	Action, Adventure, ScienceFiction, Thriller
Aliens	1.3978	Action, Horror, ScienceFiction, Thriller
Terminator 3: Rise of the Machines	1.4491	Action, ScienceFiction, Thriller
Terminator Salvation	1.4787	Action, ScienceFiction, Thriller
Surrogates	1.5352	Action, ScienceFiction, Thriller
Terminator Genisys	1.5820	Action, Adventure, ScienceFiction, Thriller
Fortress	1.6107	Action, ScienceFiction, Thriller
Doomsday	1.6377	Action, ScienceFiction, Thriller

۴- نتایج تجربی و تجزیه و تحلیل

در مسئله تحقیق، سه ماژول مختلف یک مسئولیت فردی را انجام می دهند. مدل اول یک ماژول سیستم توصیه برای یافتن فیلم های مشابه از مجموعه داده است. ماژول دوم از ورودی ماژول اول استفاده می کند و فیلم پیشنهادی را بر اساس پیش بینی محبوبیت به شش کلاس مختلف طبقه بندی می کند. علاوه بر این، ماژول سوم ماژول پیش بینی مخاطب هدف است. مطالعه تجربی هر ماژول حیاتی است زیرا سیستم توصیه فیلم‌های مشابه را برای یک فیلم معین از مجموعه داده محاسبه می‌کند. پیش‌بینی موفقیت فیلم یک مشکل چند کلاسه است. دقت ماژول پیش‌بینی ضربه فیلم به شدت به کارایی مدل اول بستگی دارد. یافتن و توصیه مشابه ترین فیلم ها برای پیش بینی کلاس فیلم در آینده ضروری بوده است. ماژول پیش‌بینی مخاطب هدف به‌طور قطعی علاقه‌مندی‌های فیلم را برای هر گروه سنی تخمین می‌زند. ما از رتبه بندی و تاریخ رای گیری همه فیلم های پیشنهادی استفاده کرده ایم.

الف) توصیه فیلم

سیستم پیشنهادی فیلم مبتنی بر محتوا از ویژگی‌هایی مانند ژانر، بازیگران، نام کارگردان، کلمات کلیدی و توضیحات فیلم برای سنجش شباهت بین دو فیلم استفاده می‌کند. شباهت بین دو فیلم پس از محاسبه فاصله شباهت از هر پارامتر محاسبه می شود.

جدول ۹. فیلم های پیشنهادی فیلم انتقام جویان.

Recommended Movie	Similarity distance	Genres
Avengers: Age of Ultron	0.5031	Action, Adventure, ScienceFiction
Ant-Man	1.0835	Action, Adventure, ScienceFiction
Captain America: The Winter Soldier	1.0908	Action, Adventure, ScienceFiction
Captain America: Civil War	1.0930	Action, Adventure, ScienceFiction
Iron Man 2	1.1864	Action, Adventure, ScienceFiction
Captain America: The First Avenger	1.2637	Action, Adventure, ScienceFiction
Iron Man	1.2887	Action, Adventure, ScienceFiction
Iron Man 3	1.3764	Action, Adventure, ScienceFiction
X-Men Origins: Wolverine	1.3784	Action, Adventure, ScienceFiction, Thriller
The Incredible Hulk	1.3816	Action, Adventure, ScienceFiction

جدول ۸ فیلم پیشنهادی فیلم "نابودگر" و جدول ۹ فیلم "انتقام‌جو" را نشان می‌دهد. جداول ۱۰ فیلم برتر مشابه هر فیلم انتخابی را ارائه کرده و فاصله شباهت کلی از فیلم انتخابی تا هر فیلم پیشنهادی را محاسبه کرده است. همچنین ژانر هر یک از فیلم های پیشنهادی را ارائه کرده ایم. سیستم توصیه فاصله شباهت را از فیلم انتخابی محاسبه کرد. فاصله شباهت بین دو فیلم، شباهت بین آنها را نشان می دهد. با کاهش فاصله، شباهت بین دو فیلم بیشتر می شود و بالعکس. بر اساس محاسبات، "ترمیناتور ۲" شبیه ترین فیلم به "نابودگر" است. فاصله تشابه کلی اندازه گیری شده ۱.۲۶۹۸ است که حداقل بین تمام فاصله ها است. به همین ترتیب، «انتقام‌جویان: عصر اولترون» شبیه‌ترین فیلم «انتقام‌جو» در مقایسه با فیلم‌های دیگر است. فاصله تشابه کلی اندازه گیری شده توسط سیستم توصیه ۰.۵۰۳۱ است. ژانرهای فیلم "نابودگر" اکشن، هیجان انگیز و علمی تخیلی هستند. همه فیلم‌های پیشنهادی منتخب نیز تقریباً ژانرهای مشابهی دارند. فیلم "انتقام جویان" دارای ژانرهای علمی تخیلی، اکشن و ماجراجویی است. همه فیلم‌های مشابه منتخب نیز تقریباً ژانرهای مشابهی دارند. جدول ۱۰ فیلم های توصیه شده برای فیلم زیبایی آمریکایی را نشان می دهد. ژانر فیلم درام است. همه فیلم‌های پیشنهادی را که ژانر مشابهی دارند نیز بشنوید. ظروف محاسبه شده کلی نیز در جدول نشان داده شده است. طبق سیستم، جاده انقلابی شبیه ترین فیلم به فیلم منتخب است و فاصله شباهت کلی ۱.۵۰۷۶ است.

ب) پیش بینی ضربه فیلم

در این بخش، پیش‌بینی موفقیت فیلم مورد بحث قرار می‌گیرد. پیش‌بینی ضربه فیلم یک مشکل طبقه‌بندی چند طبقه است. ما مشکل را به شش کلاس محبوبیت {SDH، SH، H، AA، A، F} طبقه بندی می کنیم. ماژول سیستم توصیه اولیه N نوع مشابه از فیلم های فیلم آینده را ارائه می دهد. این سیستم از پارامترهای رای گیری و رتبه بندی همه فیلم های توصیه شده استفاده می کند. در نهایت، محبوبیت فیلم آینده را پیش‌بینی می‌کند، دقت مدل پیش‌بینی اساساً به دقت مدل سیستم توصیه مربوط می‌شود.

جدول ۱۰. فیلم های پیشنهادی فیلم American Beauty.

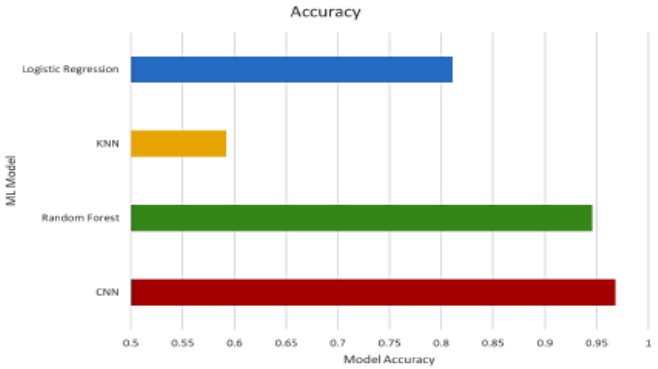
Recommended Movie	Similarity distance	Genres
Revolutionary Road	1.5076	Drama, Romance
Jarhead	1.6959	Drama, War
Whip It	1.7166	Drama
Philomena	1.7473	Drama
The Brown Bunny	1.7473	Drama
Life as a House	1.7631	Drama
The Color Purple	1.7664	Drama
My Week with Marilyn	1.7751	Drama
The Ice Storm	1.7849	Drama
Draft Day	1.7888	Drama

مجموعه داده دارای ۳۳۱۰ داده فیلم است. ما مجموعه داده ها را به نسبت ۸۰:۲۰ جدا کرده ایم. ۸۰٪ به معنای ۲۶۸۴ داده فیلم برای اهداف آموزشی و ۶۲۶ داده باقی مانده برای اهداف آزمایشی است. ما این مجموعه داده جدید را داریم و مدل‌های مختلف یادگیری ماشین را برای دقت بهتر آزمایش کرده‌ایم. جدول ۱۱ مدل های مختلف یادگیری ماشین را با استفاده از چندین پارامتر تکامل مقایسه می کند. مدل CNN پیشنهادی به طور قابل توجهی بهتر از همه مدل‌های یادگیری ماشین عمل می‌کند و تمام پارامترها را با هم مقایسه می‌کند. شکل ۸ دقت مدل یادگیری ماشین را نشان می دهد.

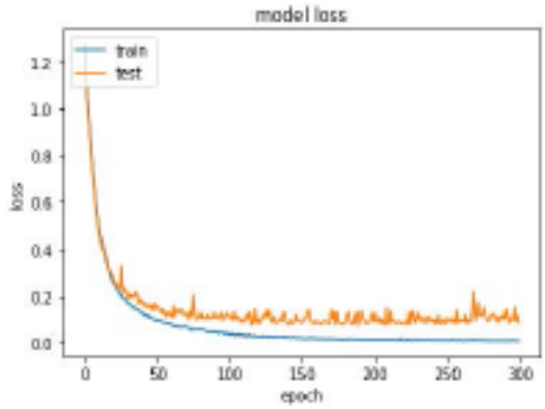
جدول ۱۱. مقایسه مدل های مختلف یادگیری ماشین

Machine Learning Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.811	0.81	0.81	0.79
KNN	0.592	0.60	0.59	0.59
Random Forest	0.946	0.95	0.95	0.94
CNN	0.968	0.96	0.96	0.96

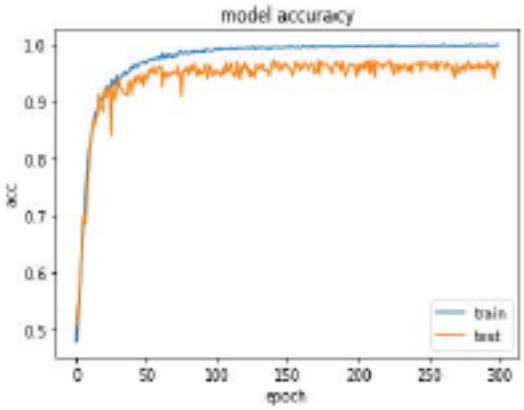
تجزیه و تحلیل تجربی مجموعه داده‌ها نشان می‌دهد که عملکرد کلی مدل شبکه عصبی پیچشی پیشنهادی از همه مدل‌های یادگیری ماشین معمولی پایه بهتر است. شکل ۹ و ۱۰ به ترتیب منحنی دقت مدل CNN و منحنی تلفات پیشنهادی را نشان می دهد.



شکل ۸. دقت مدل های مختلف یادگیری ماشین را مقایسه کنید.



شکل ۹. منحنی دقت مدل CNN.



شکل ۱۰. منحنی از دست دادن مدل CNN.

یک مطالعه تطبیقی از کار ما با برخی از کارهای تحقیقاتی مرتبط انجام شده در گذشته توسط احمد و همکاران (۲۰۱۹) [۲۲] ارائه شده که در آن از یک سیستم رأی گیری ترکیبی برای به دست آوردن پیش بینی مرحله اولیه تولید استفاده نموده اند. در این مطالعه، آنها ویژگی های جدیدی را برای بهبود کارایی پیش بینی معرفی کردند. آنها فیلم را در هشت کلاس مختلف موفقیت طبقه بندی کردند و دقت ۸۵ درصدی را به دست آوردند. عابدی و همکاران (۲۰۲۰) [۲۷] ویژگی های هر فیلم را بررسی کرد و تمام ویژگی های مربوط به مراحل اولیه پیش بینی فیلم را انتخاب کرد. آنها پنج مدل مختلف یادگیری ماشین را با طبقه بندی باینری اجرا کرده اند و با مدل خطی تعمیم یافته (GLM) به دقت ۷۶.۶ درصدی دست یافته اند. مایکل و کانگ (۲۰۱۶) [۱۰] ویژگی های جدیدی را معرفی کردند و موفقیت فیلم را در مراحل اولیه تولید فیلم پیش بینی کردند. آنها مدل های یادگیری ماشین را چندین بار با معیارهای موفقیت مشخص ارزیابی کردند. مایکل و کانگ (۲۰۱۶) با استفاده از مدل جنگل تصادفی با استفاده از طبقه بندی باینری به حداکثر دقت ۹۰.۴ درصد دست یافتند. برای طبقه بندی چند کلاسه، آنها ۸۴.۷٪ دقت را به دست آورده اند. (Verma and Garima 2019) [۲۴] «امتیاز موسیقی» را به عنوان یکی از ویژگی‌های مهم برای پیش‌بینی موفقیت‌های فیلم پیشنهاد کردند. آنها با استفاده از مدل جنگل تصادفی با استفاده از طبقه بندی باینری به دقت ۸۷.۰٪ دست یافتند. مدل پیشنهادی ما از همه مدل‌های قبلی بهتر عمل می‌کند و به دقت ۹۶.۸ درصد دست یافته است. جدول ۱۲ تجزیه و تحلیل مقایسه ای را نشان می دهد.

جدول ۱۲. تحلیل تطبیقی کار پیشنهادی ما با آثار مرتبط.

Author and Year	Proposed Work	Methodologies/ Parameters	Classification Type	Accuracy
Ahmed et al. 2019	Pre-production box-office success prediction	Incorporate new features and Hybrid voting classifier	Multiclass	0.85
Abidi et al. 2020	Movie's popularity prediction	Incorporate novel features and generalized linear model (GLM)	Binary class	0.766
Michael and Kang 2016	Movie profitability using what, who and when features.	Incorporate novel features and using RF machine learning model	Binary class	0.904
Michael and Kang 2016	Movie profitability using what, who and when features.	Incorporate novel features and using RF machine learning model	Multiclass	0.847
Verma and Garima 2019	Bollywood movie success prediction	Incorporate music rating and use RF machine learning model	Binary class	0.87
Our proposed system	Predicting movie success	Using content-based RS and using CNN deep learning model	Multiclass	0.968

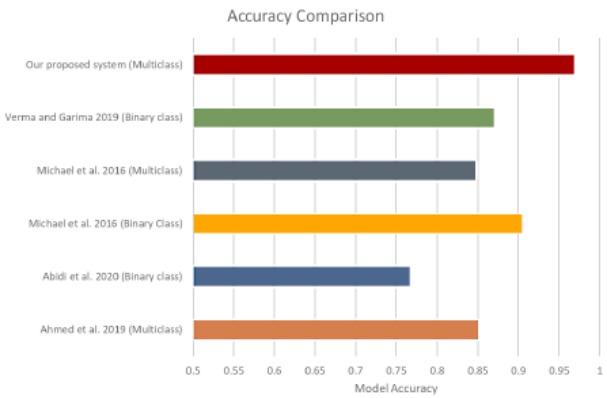
ج) پیش بینی گروه مخاطب ترجیحی

این بخش فرعی نتیجه و تجزیه و تحلیل ماژول نهایی پیش بینی مخاطب هدف فیلم را مورد بحث قرار می دهد. مخاطبان در چهار گروه سنی (جوانان، نوجوانان، میانسالان، سالمندان) گروه بندی می شوند. برای پیش‌بینی مخاطبان فیلم آینده، ما همه فیلم‌های پیشنهادی ارائه‌شده توسط اولین ماژول خود را در نظر گرفته‌ایم. سپس، سیستم از رتبه بندی هر فیلم توصیه شده توسط هر گروه سنی استفاده می کند و تعداد جزئیات رای هر گروه را برای تجزیه و تحلیل سیستم فیلم توصیه شده از تمام اطلاعات رای و رتبه بندی از هر گروه از همه فیلم های توصیه شده می گیرد. در نهایت، این ماژول محبوبیت فیلم آینده را برای هر گروه سنی پیش بینی می کند.

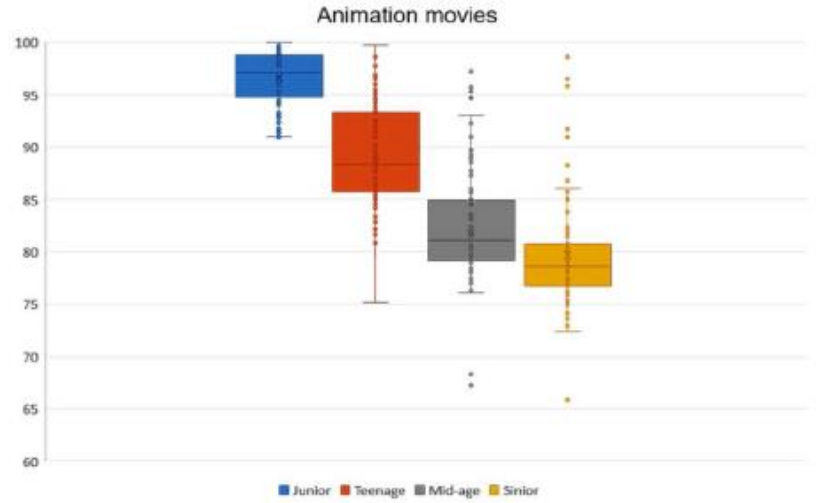
جدول ۱۳. مخاطب هدف فیلم انیمیشن

Movie title	Genres	Popularity			
		Junior	Teenage	Mid-age	Senior
Cars	Animation, Adventure, Comedy, Family	99.71	95.34	89.76	88.57
The Good Dinosaur	Adventure, Animation, Family	95.91	87.24	78.92	75.91
WALLÂ·E	Animation, Family	96.24	89.13	81.19	78.62
Kung Fu Panda	Adventure, Animation, Family, Comedy	98.83	89.7	81.4	78.49
Ice Age: Dawn of the Dinosaurs	Animation, Comedy, Family, Adventure	91.92	83.1	78.03	77.15

برخی از فیلم ها فقط گروه سنی خردسال را هدف قرار می دهند. به طور کلی، فیلم های انیمیشن به طور خاص برای گروه سنی خردسال هدف قرار می گیرند. سیستم پیشنهادی ما بر میزان محبوبیت فیلم آینده در بین همه گروه‌های سنی متمرکز بود. سازندگان فیلم می توانند محبوبیت انیمیشن آینده را در میان رده سنی خردسال تخمین بزنند. معمولاً فیلم‌های پرطرفدار همیشه تأثیر خوبی در بین تمام گروه‌های سنی دارند. جدول ۱۳ محبوبیت پنج فیلم انیمیشن پیشرو را در بین تمام گروه‌های سنی نشان می‌دهد. شکل ۱۲ محبوبیت فیلم های انیمیشن را در گروه خردسال نشان می دهد و همچنین کاهش تدریجی محبوبیت را با افزایش سن نشان می دهد.



شکل ۱۱. تحلیل تطبیقی کار پیشنهادی ما با آثار مرتبط.



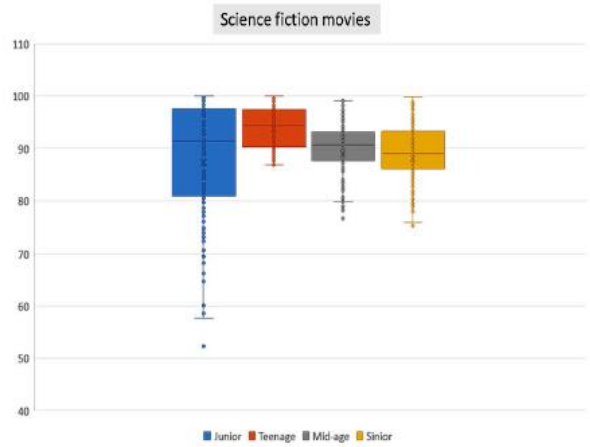
شکل ۱۲. محبوبیت طرح جعبه فیلم های انیمیشن.

گروه های سنی بالاتر معمولاً فیلم هایی مانند کمدی، درام و عاشقانه را ترجیح می دهند. فیلمسازان معمولاً گروه های میانسال و ارشد را برای ژانرهای کمدی و درام برای فیلم ها هدف قرار می دهند. جدول ۱۴ محبوبیت پنج فیلم از ژانرهای کمدی و درام را در بین تمام گروه های سنی سالمندان و میانسالان نشان می دهد که این گونه فیلم ها را بیشتر در بین تمام گروه های سنی دوست دارند. شکل ۱۳ محبوبیت فیلم های کمدی را به استثنای فیلم های علمی تخیلی، ماجراجویی و انیمیشن نشان می دهد. این نشان می دهد که محبوبیت در میان گروه سالمندان حداکثر است، گروه میانسالی بعدی. محبوبیت در میان جوانان کمترین میزان است. فیلم‌های علمی تخیلی معمولاً برای همه گروه‌های سنی ترجیح داده می‌شوند، بسته به خط داستانی فیلم، به‌ویژه گروه نوجوانان که بیشتر فیلم‌های علمی تخیلی هستند. جدول ۱۵ محبوبیت پنج فیلم علمی تخیلی پیشرو را در بین تمام گروه‌های سنی نشان می‌دهد که گروه نوجوان بیشترین علاقه را در بین تمام گروه‌های سنی دارند. هر پنج فیلم فیلم های موفق هستند. از این رو آنها در بین تمام گروه های سنی محبوب هستند. شکل ۱۴ میزان محبوبیت فیلم های علمی تخیلی را در بین تمامی گروه ها نشان می دهد. این نشان می دهد که محبوبیت در بین گروه نوجوان حداکثر است. محبوبیت در گروه جوان تا حد زیادی متفاوت است. میانگین محبوبیت فیلم های علمی تخیلی معمولا بالاست.

Movie title	Genres	Popularity			
		Junior	Teenage	Mid-age	Senior
Eulogy	Comedy, Drama	54.8	69.53	70.95	73.3
Death at a Funeral	Comedy, Drama	52.3	67.8	70.68	73.22
And So It Goes	Comedy, Drama, Romance	63.45	83.27	85.33	88.06

جدول ۱۴. مخاطبین فیلم کمدی

شکل ۱۳. محبوبیت طرح جعبه فیلم های کمدی شامل فیلم های علمی تخیلی، ماجراجویی و انیمیشن نمی شود.



شکل ۱۴. محبوبیت طرح جعبه فیلم های علمی تخیلی.

۵- نتیجه گیری

مقدار قابل توجهی از بودجه در هر فیلم باکس آفیس مصرف می شود. با این حال، بیشتر فیلم ها موفق نمی شوند. پیش از این، بیشترین تعداد کار در زمینه پیش‌بینی پس از تولید یا پس از انتشار انجام شده است. این برآورد تأثیری ندارد زیرا سرمایه گذار قبلاً سرمایه خود را برای تولید فیلم مصرف کرده است. پیش‌بینی مرحله پیش‌تولید یا تولید اولیه به دقت بالا و بهترین زمان برای اطمینان از سرمایه‌گذاری نیاز دارد. هدف از مطالعه ما پیشنهاد یک سیستم خبره است که بتواند به سازنده فیلم کمک کند تا در صورت نیاز تغییرات لازم را در زمان مناسب انجام دهد. سیستم ما می تواند قبل از شروع تولید برای اولین مرحله تولید و با دقت قابل توجهی، به میزان محبوبیت فیلم آینده هزینه کند. درباره سیستم نه تنها بر محبوبیت فیلم آینده بلکه بر محبوبیت فیلم در بین تمام گروه‌های سنی تمرکز داشت. Movie Maker می‌تواند مخاطب هدف را تخمین بزند و ارزیابی کند که گروه‌های مختلف مخاطب چگونه به فیلم آینده پاسخ خواهند داد. علاوه بر این، هدف ما ساختن یک سیستم قوی است که برای همه صنایع فیلمسازی قابل اجرا باشد. ما از صد سال گذشته (۱۹۱۵-۲۰۱۶) داده های فیلم از TMDband پایگاه داده IMDb استفاده کرده ایم. رویکرد ما به محبوبیت فیلم متمرکز و پیدا کردن مخاطب هدف یک فیلم آینده بسیار منحصر به فرد است. در رویکرد خود، ما از یک سیستم توصیه برای یافتن فیلم‌های مشابه از یک فیلم خاص و استفاده از فیلم‌های مشابه برای اهداف پیش‌بینی استفاده کرده‌ایم. علاوه بر این، استفاده همزمان برای جداسازی پایگاه داده (TMDb و IMDb) چالش برانگیز بوده است. اندازه مجموعه داده TMDb 4803 است و اندازه مجموعه داده رتبه بندی IMDb 85855 است. از آنجایی که ما از هر دو مجموعه داده استفاده می کنیم، اندازه مجموعه داده ادغام شده تنها به ۲۳۳۳۲ کاهش می یابد. ما باید در مورد هر یک از

ویژگی هایی که می تواند در ابتدای تولید فیلم در دسترس باشد قضاوت کنیم. ما فقط پنج ویژگی فیلم را برای سیستم توصیه خود به دقت انتخاب کرده‌ایم. از مجموع آرا برای هر فیلم استفاده کرده‌ایم. مشاهده شده است که رای فیلم قدیمی نسبتا کمتر از فیلم جدید است. علاوه بر این، چندین اطلاعات رای گیری برای گروه خردسال به طور قابل توجهی نسبت به سایر گروه ها کمتر است. استفاده از اطلاعات رای‌گیری و رتبه‌بندی برای ایجاد یک مجموعه ویژگی جدید برای کار در این سناریو چالش برانگیز است. سیستم پیشنهادی یک ابزار عالی برای صنعت فیلم است. در کارهای آینده، داده های چند رسانه ای مانند داده های صوتی و تصویری می تواند گنجانده شود و همچنین می توان از پوستر فیلم آینده برای نتایج بهتر استفاده کرد. بلیط های قطار اخیر را می توان با استفاده از تحلیل احساسات داده های رسانه های اجتماعی تجزیه و تحلیل کرد. اطلاعات مربوط به آموزش های اخیر در مورد انتظارات بازار از صنعت فیلم برای سازندگان فیلم مفید خواهد بود. گروه مخاطب را می توان بر اساس سن و بر اساس جمعیت شناسی یا حرفه مخاطب تقسیم کرد. این برای هدف قرار دادن و تبلیغ یک فیلم آینده بسیار آسان تر خواهد بود.

منابع

[1] L. Sharma and A. Gera, ``A survey of recommendation system: Research challenges," Int. J. Eng. Trends Technol., vol. 4, no. 5, pp. 1989_1992, 2013.
[2] N. Das, S. Borra, N. Dey, and S. Borah, ``Social networking in web-based movie recommendation system," in Social Networks Science: Design, Implementation, Security, and Challenges. Cham, Switzerland: Springer, 2018, pp. 25_45.
[3] P. Nagarnaik and A. Thomas, ``Survey on recommendation system methods," in Proc. 2nd Int. Conf. Electron. Commun. Syst. (ICECS), Feb. 2015, pp. 1603_1608.
[4] M. A. Hameed, O. Al Jadaan, and S. Ramachandram, ``Collaborative _ltering based recommendation system: A survey," Int. J. Comput. Sci. Eng., vol. 4, no. 5, p. 859, 2012.
[5] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, ``Item-based collaborative_ltering recommendation algorithms," in Proc. 10th Int. Conf.WorldWide Web (WWW), 2001, pp. 285_295.
[6] Y. Koren and R. Bell, ``Advances in collaborative_ltering," in Recommender Systems Handbook. Boston, MA, USA: Springer, 2015, pp. 77_118.
[7] J. L. Herlocker, J. A. Konstan, and J. Riedl, ``Explaining collaborative _ltering recommendations," in Proc. ACM Conf. Comput. supported Coop-erat. Work (CSCW), 2000, pp. 241_250.
[8] M. J. Pazzani, ``A framework for collaborative, content-based and demographic _ltering," Artif. Intell. Rev., vol. 13, no. 5, pp. 393_408, 1999.
[9] R. Van Meteren and M. Van Someren, ``Using content-based _ltering for recommendation," in Proc. Mach. Learn. Inf. Age, MLnet/ECML2000 Workshop, vol. 30, 2000, pp. 47_56.
[10] P. B. Thorat, R. M. Goudar, and S. Barve, ``Survey on collaborative_ltering, content-based _ltering and hybrid recommendation system," Int. J. Comput. Appl., vol. 110, no. 4, pp. 31_36, Jan. 2015.
[11] P. Lops, M. Gemmis, and G. Semeraro, ``Content-based recommender systems: State of the art and trends," in Recommender Systems Handbook. Boston, MA, USA: Springer, 2011, pp. 73_105.
[12] M. J. Pazzani and D. Billsus, ``Content-based recommendation systems," in The Adaptive Web. Berlin, Germany: Springer, 2007, pp. 325_341.
[13] D. Wang, Y. Liang, D. Xu, X. Feng, and R. Guan, ``A content-based recommender system for computer science publications," Knowl.-Based Syst., vol. 157, pp. 1_9, Oct. 2018.
[14] R. Burke, ``Hybrid recommender systems: Survey and experiments," User Modeling and User-Adapted Interaction, vol. 12, no. 4, pp. 331_370, 2002.
[15] M. A. Ghazanfar and A. Prugel-Bennett, ``A scalable, accurate hybrid recommender system," in Proc. 3rd Int. Conf. Knowl. Discovery Data Mining, Jan. 2010, pp. 94_98.
[16] T. K. Paradarami, N. D. Bastian, and J. L. Wightman, ``A hybrid recommender system using arti_cial neural networks," Expert Syst. Appl., vol. 83, pp. 300_313, Oct. 2017.
[17] X. Su and T. M. Khoshgoftaar, ``A survey of collaborative _ltering techniques," in Advances in Arti_cial Intelligence. London, U.K., 2009.
[18] P. Melville and V. Sindhwani, ``Recommender systems," Encyclopedia Mach. Learn., vol. 1, pp. 829_838, Apr. 2010.
[19] Y. Yao and F. M. Harper, ``Judging similarity: A user-centric study of related item recommendations," in Proc. 12th ACM Conf. Recommender Syst., Sep. 2018, pp. 288_296.
[20] F. Gedikli, D. Jannach, and M. Ge, ``How should i explain? A comparison of different explanation types for recommender systems," Int. J. Hum.-Comput. Stud., vol. 72, no. 4, pp. 367_382, Apr. 2014.
[21] N. J. Belkin and W. B. Croft, ``Information _ltering and information retrieval: Two sides of the same coin?" Commun. ACM, vol. 35, no. 12, pp. 29_38, Dec. 1992.
[22] T. Di Noia, R. Mirizzi, V. C. Ostuni, D. Romito, and M. Zanker, ``Linked open data to support content-based recommender systems," in Proc. 8th Int. Conf. Semantic Syst. (SEMANTICS), 2012, pp. 1_8.
[23] C. Musto, P. Lops, M. de Gemmis, and G. Semeraro, ``Semantics-aware recommender systems exploiting linked open data and graph-based features," Knowl.-Based Syst., vol. 136, pp. 1_14, Nov. 2017.
[24] M. Uluyagmur, Z. Cataltepe, and E. Tayfur, ``Content-based movie recommendation using different feature sets," in Proc. World Congr. Eng. Comput. Sci., vol. 1, 2012, pp. 17_24.
[25] S. R. S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, ``Content-based movie recommendation system using genre correlation," in Smart Intelligent Computing and Applications. Singapore: Springer, 2019, pp. 391_397.

[26] J. Son and S. B. Kim, ``Content-based _ltering for recommendation systems using multiattribute networks," Expert Syst. Appl., vol. 89, pp. 404_412, Dec. 2017.
[27] S. M. Ali, G. K. Nayak, R. K. Lenka, and R. K. Barik, ``Movie recommendation system using genome tags and content-based _ltering," in Advances in Data and Information Sciences. Singapore: Springer, 2018, pp. 85_94.
[28] M. Elahi, Y. Deldjoo, F. Bakhshandegan Moghaddam, L. Cella, S. Cereda, and P. Cremonesi, ``Exploring the semantic gap for movie recommendations," in Proc. 11th ACM Conf. Recommender Syst., Aug. 2017, pp. 326_330.
[29] Y. Deldjoo, M. F. Dacrema, M. G. Constantin, H. Eghbal-zadeh, S. Cereda, M. Schedl, B. Ionescu, and P. Cremonesi, ``Movie genome: Alleviating newitem cold start in movie recommendation," User Model. User-Adapted Interact., vol. 29, no. 2, pp. 291_343, Apr. 2019.
[30] S. Kumar, K. De, and P. P. Roy, ``Movie recommendation system using sentiment analysis from microblogging data," IEEE Trans. Computat. Social Syst., vol. 7, no. 4, pp. 915_923, Aug. 2020.
[31] S. Philip, P. B. Shola, and A. Ovye, ``Application of content-based approach in research paper recommendation system for a digital library," Int. J. Adv. Comput. Sci. Appl., vol. 5, no. 10, pp. 1_4, 2014.
[32] J. Eliashberg, J.-J. Jonker, M. S. Sawhney, and B. Wierenga, ``MOVIEMOD: An implementable decision-support system for prerelease market evaluation of motion pictures," Marketing Sci., vol. 19, no. 3, pp. 226_243, Aug. 2000.
[33] P. Boccardelli, F. Brunetta, and F. Vicentini, ``What is critical to success in the movie industry? A study on key success factors in the Italian motion picture industry," Tech. Rep., 2008. [Online]. Available: https://publires.unicatt.it/en/publications/what-is-critical-to-successin- the-movie-industry-a-study-on-key_10.
[34] W. Zhang and S. Skiena, ``Improving movie gross prediction through news analysis," in Proc. IEEE/WIC/ACMInt. Joint Conf.Web Intell. Intell. Agent Technol., Sep. 2009, pp. 301_304.
[35] W. Bai, Y. Zhang,W. Huang, Y. Zhou, D.Wu, G. Liu, and L. Xiao, ``Deep-Fusion: Predicting movie popularity via cross-platform feature fusion," Multimedia Tools Appl., vol. 79, nos. 27_28, pp. 19289_19306, Jul. 2020.
[36] M. Galvão and R. Henriques, ``Forecasting movie box of_ce pro_tability," J. Inf. Syst. Eng. Manage., vol. 3, no. 3, pp. 1_9, Jul. 2018.
[37] Y. Zhou, L. Zhang, and Z. Yi, ``Predicting movie box-of_ce revenues using deep neural networks," Neural Comput. Appl., vol. 31, no. 6, pp. 1855_1865, Jun. 2019.
[38] K. Lee, J. Park, I. Kim, and Y. Choi, ``Predicting movie success with machine learning techniques: Ways to improve accuracy," Inf. Syst. Fron-tiers, vol. 20, no. 3, pp. 577_588, Jun. 2018.
[39] M. T. Lash and K. Zhao, ``Early predictions of movie success: The who, what, and when of pro_tability," J. Manage. Inf. Syst., vol. 33, no. 3, pp. 874_903, Jul. 2016.
[40] U. Ahmed, H. Waqas, and M. T. Afzal, ``Pre-production box-of_ce success quotient forecasting," Soft Comput., vol. 24, no. 9, pp. 6635_6653, May 2020.
[41] S. M. R. Abidi, Y. Xu, J. Ni, X. Wang, and W. Zhang, ``Popularity prediction of movies: From statistical modeling to machine learning techniques," Multimedia Tools Appl., vol. 79, nos. 47_48, pp. 35583_35617, Dec. 2020.
[42] H. Verma and G. Verma, ``Prediction model for bollywood movie success: A comparative analysis of performance of supervised machine learning algorithms," Rev. Socionetwork Strategies, vol. 14, no. 1, pp. 1_17, Apr. 2020.
[43] M. Mestyán, T. Yasseri, and J. Kertész, ``Early prediction of movie box of_ce success based on Wikipedia activity big data," PLoS ONE, vol. 8, no. 8, Aug. 2013, Art. no. e71226.
[44] K. R. Apala, M. Jose, S. Motnam, C.-C. Chan, K. J. Liszka, and F. de Gregorio, ``Prediction of movies box of_ce performance using social media," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Aug. 2013, pp. 1209_1214.
[45] S. Asur and B. A. Huberman, ``Predicting the future with social media," in Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol., Aug. 2010, pp. 492_499.
[46] S. Gopinath, P. K. Chintagunta, and S. Venkataraman, ``Blogs, advertising, and local-market movie box of_ce performance," Manage. Sci., vol. 59, no. 12, pp. 2635_2654, Dec. 2013.
[47] R. Parimi and D. Caragea, ``Pre-release box-of_ce success prediction for motion pictures," in Proc. Int. Workshop Mach. Learn. Data Mining Pattern Recognit. Berlin, Germany: Springer, 2013, pp. 571_585.

[48] J. S. Simonoff and I. R. Sparrow, ``Predicting movie grosses: Winners and losers, blockbusters and sleepers," Chance, vol. 13, no. 3, pp. 15_24, Jun. 2000.
[49] M. Baimbridge, ``Movie admissions and rental income: The case of James Bond," Appl. Econ. Lett., vol. 4, no. 1, pp. 57_61, Jan. 1997.
[50] A. De Vany and W. D. Walls, ``Uncertainty in the movie industry: Does star power reduce the terror of the box of_ce?" J. Cultural Econ., vol. 23, no. 4, pp. 285_318, 1999.
[51] A. Elberse, ``The power of stars: Do star actors drive the success of movies?" J. Marketing, vol. 71, no. 4, pp. 102_120, Oct. 2007.
[52] J. Eliashberg, S. K. Hui, and Z. J. Zhang, ``From story line to box of_ce: A new approach for green-lighting movie scripts," Manage. Sci., vol. 53, no. 6, pp. 881_893, Jun. 2007.
[53] W. D. Walls, ``Modeling movie success when `nobody knows anything': Conditional stable-distribution analysis of _lm returns," J. Cultural Econ., vol. 29, no. 3, pp. 177_190, Aug. 2005.
[54] A. Zaheer and G. Soda, ``Network evolution: The origins of structural holes," Administ. Sci. Quart., vol. 54, no. 1, pp. 1_31, Mar. 2009.
[55] J. Du, H. Xu, and X. Huang, ``Box of_ce prediction based on microblog," Expert Syst. Appl., vol. 41, no. 4, pp. 1680_1689, Mar. 2014.
[56] R. Sharda and D. Delen, ``Predicting box-of_ce success of motion pictures with neural networks," Expert Syst. Appl., vol. 30, no. 2, pp. 243_254, Feb. 2006.
[57] L. Zhang, J. Luo, and S. Yang, ``Forecasting box of_ce revenue of movies with BP neural network," Expert Syst. Appl., vol. 36, no. 3, pp. 6580_6587, Apr. 2009.
[58] V. R. Nithin, ``Predicting movie success based on imdb data," Int. J. Res. Appl. Sci. Eng. Technol., vol. V, no. X, pp. 504_507, Oct. 2017.
[59] Box Of_ce Revenue in the United States and Canada from 1980 to 2021. Accessed: May 5, 2021. [Online]. Available: https://www.statista.com/statistics/187069/north-american-box-of_cegross-revenue-since1980/
[60] B. Meiseberg and T. Ehrmann, ``Diversity in teams and the success of cultural products," J. Cultural Econ., vol. 37, no. 1, pp. 61_86, Feb. 2013.
[61] N. Kurniasih, E. Rizal, Y. Winoto, Sukaesih, N. Kurniawati, Sujito, A. Sudirman, A. Hasibuan, A. Daengs GS, and K. Saddhono, ``Online media as a movie reference," J. Phys., Conf., vol. 1114, Nov. 2018, Art. no. 012087.
[62] Kaggle TMDB Movie Dataset. Accessed: May 5, 2021. [Online]. Available: https://www.kaggle.com/tmdb/tmdb-movie-metadata.
[63] Kaggle IMDB Movie Dataset. Accessed: May 5, 2021. [Online]. Available: https://www.kaggle.com/stefanoleone992/imdb-extensivedataset?select=IMDb+ratings.csv
[64] M. A. Belarbi, S. Mahmoudi, and G. Belalem, ``PCA as dimensionality reduction for large-scale image retrieval systems," Int. J. Ambient Comput. Intell., vol. 8, no. 4, pp. 45_58, Oct. 2017.
[65] S. K. Das, S. P. Das, N. Dey, and A. E. Hassanien. Machine Learning Algorithms for Industrial Applications. Amsterdam, The Netherlands: Elsevier,2020.
[66] S. Sivakumar and R. Rajalakshmi, ``Analysis of sentiment on movie reviews using word embedding self-attentive LSTM," Int. J. Ambient Comput. Intell., vol. 12, no. 2, pp. 33_52, Apr. 2021.
[67] W. Chen, H.-T. Zheng, Y. Wang, W. Wang, and R. Zhang, ``Utilizing generative adversarial networks for recommendation based on ratings and reviews," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2019, pp. 1_8.
[68] Y.Wang, H.-T. Zheng,W. Chen, and R. Zhang, ``LambdaGAN: Generative adversarial nets for recommendation task with lambda strategy," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2019, pp. 1_8.