

اللّٰهُ الرَّحْمٰنُ الرَّحِيْمُ



دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

تحلیل داده و پیاده سازی سیستم پیشنهاد دهنده کتاب صوتی

استاد راهنما

دکتر ریحانه ریخته گران

گزارش پژوهی کارشناسی

فرنوش کریمی

۹۷۳۳۱۳۳

۱۴۰۱ بهمن

فهرست

5.....	مقدمه
6.....	جمع آوری و استخراج داده
10	تصویرسازی داده
10	تعداد آراء و امتیازات ۱.۲
13	ترجمه ۲.۲
14	مدت زمان ۳.۲
16	سال انتشار ۴.۲
17	نویسنده ۵.۲
18	قیمت ۶.۲
20	خلاصه کتاب ۷.۲
22	پیش پردازش
27	سیستم پیشنهاد دهنده کتاب
37	پیوست
38	منابع

مقدمه

سیستم پیشنهاد دهنده^۱ سعی می کند علاوه بر یک کاربر خاص در وبسایت شمارا شناسایی و پیش‌بینی کند و گزینه‌های متناسب با علاوه‌یافش به وی نشان دهد. سیستم پیشنهاد دهنده از کاربردهای مرسوم یادگیری ماشین^۲ است که خرده فروشی‌های آنلاین متعددی از آن برای افزایش فروش خود استفاده می‌کنند. همچنین در بخش‌های دیگر کاربردهای ویژه‌ای دارد.

در این پژوهه ابتدا داده‌های سایت کتابراه که مربوط به کتاب‌های صوتی ایرانی و خارجی است، استخراج شده و اعمالی مانند مصور سازی داده و پیاده سازی سیستم پیشنهاد دهنده انجام شده و سپس با یک رابط کاربری برای استفاده آمده شده است.

اهمیت و هدف این پژوهه، ساده‌تر شدن انتخاب کتاب صوتی مورد علاقه بر اساس کتاب‌های خوانده شده، توسط یک سیستم پیشنهاد دهنده کتاب و بهتر در ک کردن برخی روابط بین عوامل مرتبط با این موضوع است. این پژوهه از پنج بخش اصلی تشکیل شده است.

۱. در قسمت اول، داده‌های مربوط به پژوهه، از طریق استخراج داده از وب^۳ از سایت کتابراه و با استفاده از نرم افزارهای مربوط بصورت دستی استخراج شده است

۲. در قسمت دوم، برای استخراج مفاهیم و در ک بهتر داده (تحلیل داده‌ها و بی بردن به برخی روابط بین داده‌ها) از مصورسازی داده^۴ به وسیله نرم افزار پاور بی آی^۵ استفاده شده است

۳. در قسمت سوم، برای آماده سازی و مرتب سازی داده برای ورود به الگوریتم از تکنیک‌های پیش پردازش داده^۶ استفاده شده است

۴. در قسمت چهارم، سیستم پیشنهاد دهنده کتاب برای پیشنهاد کتاب‌های مشابه با نام کتاب وارد شده توسط کاربر، پیاده سازی شده است

۵. در قسمت پایانی، سیستم پیشنهاد دهنده به عنوان یک نرم افزار با رابط کاربری ارائه شده است

در ادامه به تفسیر مفصل این بخش‌ها خواهیم پرداخت.

¹ Recommender System

² Machine Learning

³ Web Scraping

⁴ Data Visualization

⁵ Power Bi

⁶ Data Preprocessing

فصل اول

جمع آوری و استخراج داده

در ابتدا برای جمع آوری داده‌ی بروز ایرانی برای کتاب‌های صوتی به چند منبع معتبر رجوع شد اما داده‌ای که دارای همه‌ی پویزگی‌های مورد نظر باشد (تعداد بالا، تکمیل بودن و ...) در سطح اینترنت موجود نبود. بنابراین از طریق کتابخانه‌ی بیوتفول سوب^۷ زبان پایتون^۸ و نرم افزار استخراج داده اکتوپارس^۹، داده‌ها بصورت دستی و دلخواه گردآوری شد.

کتابراه[1] مرجع قانونی دانلود کتاب الکترونیکی و دانلود کتاب صوتی است که امکان دسترسی به هزاران کتاب، رمان، مجله و کتاب صوتی و همچنین خرید کتاب الکترونیک از طریق موبایل تبلت و رایانه برای شما فراهم می‌کند و داده‌ها به شکل زیر است:

Author	Translator	Speaker	Publisher	Year	Format	Time	Price	Genre	Votes	Translated	Images	About	Name
پالولو کونیلو		احمد گنجی	حسین نعیی	نشر ثالث	۱۳۹۶	MP3	۷ ساعت و ۴۱ دقیقه	۳۴.۰۰ ت - ۶.۴۹ یورو	(۳۱۲۲ رای) کلاسیک ادبیات	1.0	https://img.ketabrah.ir/img/I/8153564537814758...	پسرک «سانشینگ» کتاب ناداشت. هواگرد کیمیکر و میش بود که...	
الیف شافاک		احمد گنجی، علی نبوی، ساروی، شهاب درخشش، ... ارمن		انتشارات فتوس	۱۳۹۷	MP3	۱۹ ساعت و ۵۵ دقیقه	۵۳.۰۰ ت - ۹.۹۹ یورو	(۳۲۲۳ رای) خارجی عاشقانه	1.0	https://img.ketabrah.ir/img/I/6701643057778123...	کتاب دلیا چاه پریشانی است نبود شمس. عشقی از پس رفتش (اجرام) ...	
اروین یاوم		سلطان زاده	سیده حبیب	نشر قطره	۱۳۹۷	MP3	۱۶ ساعت و ۴۹ دقیقه	۵۱.۰۰ ت - ۹.۴۹ یورو	(۱۷۰۵ رای) خارجی روانشناسی	1.0	https://img.ketabrah.ir/img/I/8558201548331221...	هر - کتاب انسانی، میلک مرگ خود نجه و میتواند به ر...	
مت هیگ		فاطمه فراهانی	محمد صالح نورانی زاده	انتشارات کتاب کله پُشنی	۱۴۰۰	MP3	۱۰ ساعت و ۲۸ دقیقه	۲۸.۰۰ ت - ۵.۹۹ یورو	(۴۵۷ رای) علمی و تخلیلی خارجی	1.0	https://img.ketabrah.ir/img/I/8574475464041519...	حرمتها ادعاوار کتابخانه نیمه شب...	
انتوان دو سنت اگزوپری		میثم رحمتی	میثم رحمتی	انتشارات پر عسگری	۱۳۹۸	MP3	۲ ساعت و ۱ دقیقه	۹.۰۰ ت - ۱.۹۹ یورو	(۹۵۷ رای) کلاسیک ادبیات	1.0	https://img.ketabrah.ir/img/I/6790259418906252...	رویداد گفت تو هنوز برای من شازده پرسنجمای کرجولو بیش نیست	

شکل ۱

^۷Beautiful Soup

^۸Python

^۹Octoparse

تعداد کتاب‌ها ۱۰۲۱ عدد و متغیرهای توضیحی به شکل زیر هستند:

<i>Author</i>	نام نویسنده
<i>Translator</i>	مترجم
<i>Speaker</i>	گوینده
<i>Publisher</i>	انتشارات
<i>Year</i>	سال انتشار
<i>Format</i>	-
<i>Duration</i>	مدت زمان کتاب صوتی (دقیقه)
<i>Price</i>	قیمت (تومان)
<i>Genre</i>	ژانر
<i>Votes</i>	تعداد رای
<i>Translated</i>	کتاب ترجمه شده است یا خیر؟
<i>Images</i>	پوستر کتاب
<i>About</i>	مختصری از کتاب
<i>Name</i>	نام کتاب
<i>Ranking</i>	امتیاز کتاب از ۵

نام نویسنده همان نویسنده‌های کتاب‌ها هستند که به دو دسته ایرانی و غیر ایرانی تقسیم شده‌اند. همانطور که انتظار می‌رود ممکن است یک نویسنده چندین کتاب را نویسنده‌گی کرده باشد.

ستون مترجم‌ها فقط متعلق به کتاب‌های غیر ایرانی است و زبان کتاب‌های ترجمه نشده از انگلیسی، اسپانیایی و ... به زبان فارسی است.

ستون گوینده نیز نمایشگر صدایشگان کتاب‌ها که در برخی موارد حاصل همکاری چند گوینده است. قیمت‌ها مربوط به کتاب‌های صوتی در سایت کتابراه است که در برخی موارد با تخفیف ثبت شده‌اند.

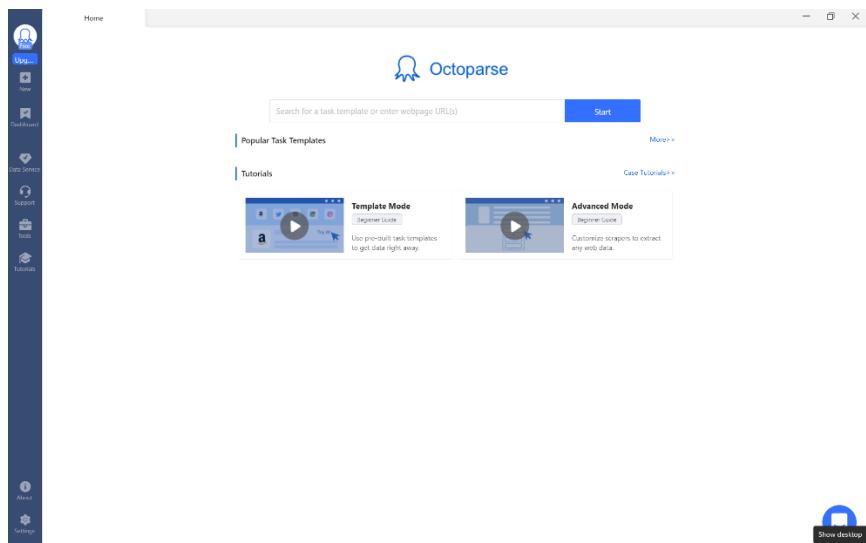
ژانر کتاب‌ها شامل بخش‌های اجتماعی، تاریخی، ترسناک، جنایی و معماهی، روانشناسی، طنز، عاشقانه، علمی تخیلی، فلسفی، مذهبی، واقعی، ادبیات کلاسیک، ادبیات مدرن، داستان کهن فارسی، داستان کوتاه و نمایش صوتی است که هر کدام دو نوع فارسی و خارجی دارند.

تعداد آراء، تعدادیست که در سایت کتابراه کاربرانی که ثبت نام کاربری شده‌اند ثبت کرده‌اند.

ستون مختصری از کتاب، مرتبط با متنی کوتاه چکیده از کتاب است که باعث آشنایی مخاطب با کتاب می‌شود.

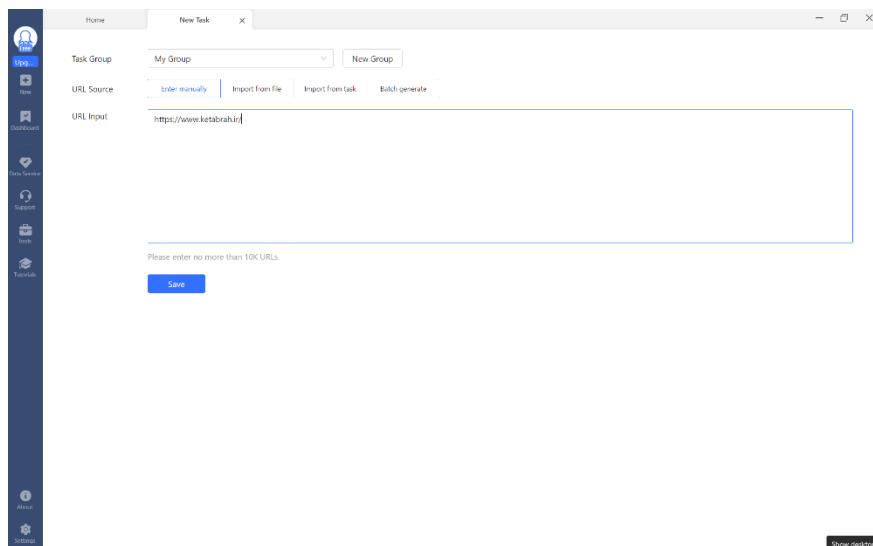
امتیاز کتاب‌ها از ۵ است که البته اکثر کتاب‌ها از بخش پرطرفدارها انتخاب و استخراج شده‌اند که امتیاز بالایی دارند.

آشنایی با محیط اکتوپارس:



شکل ۱

اکتوپارس [2] یکی از ابزار بسیار قدرتمند برای استخراج داده می‌باشد که برای گردآوری داده‌ها از این نرم افزار استفاده شده. در مرحله اول آدرس وبسایت^{۱۰} مورد نظر را وارد می‌کنیم یا اگر داده‌ها بصورت فایل یا انواع دیگری بودند با توجه به راهنمای این نرم افزار استفاده می‌توان کرد. در این قسمت آدرس سایت کتابراه را در قسمت تازه‌ها^{۱۱} و در بخش وارد کردن بصورت دستی^{۱۲} لینک را وارد می‌کنیم.



شکل ۲

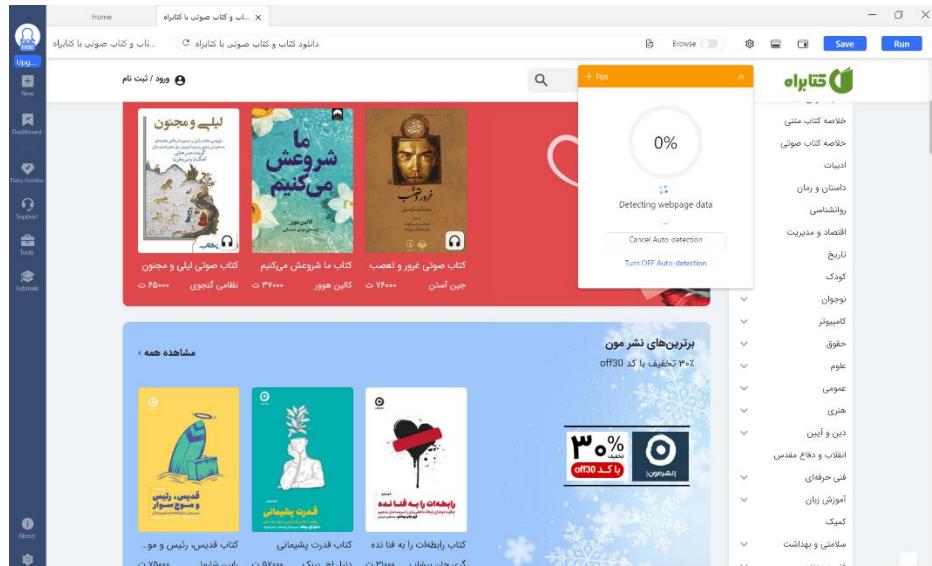
سپس وارد صفحه مورد نظر که می‌خواهیم دیتا را از آن استخراج کنیم می‌شویم و هر اطلاعی نیاز داریم آن را فقط برای داده اول انتخاب می‌کنیم و سپس تعداد داده موردنظر را وارد کرده و این نرم افزار بصورت خودکار برای داده‌های بعدی اطلاعات را استخراج می‌کند و

¹⁰ URL Address

¹¹ New File

¹² Enter Manually

داده‌های جمع آوری شده در این نرم افزار را می‌توان با فرمتهای مختلفی مانند اکسل، ورد... خروجی گرفت و ویرایش‌های لازم را در دیتابست پایه سازی کرد.



شکل ۶

فصل دوم

مصورسازی داده

برای در ک بهتر برخی الگوهای در دادهها به تحلیل از طریق مصورسازی می پردازیم.

Speaker	Ranking	Author	Translated	Publisher	Genre	Images
یوسف علیخانی	4.20	یوناس یوناسن	1	واوچوان	اجتماعی ایرانی	
یوسف علیخانی	4.60	پوکو میشمیا	1	نوین کتاب گویا	اجتماعی خارجی	
یاشار ابراهیمی	4.00	پوکو لوگاو	1	نشر ورد و دا	تاریخی ایرانی	
یاشار ابراهیمی	4.10	پوسفلی میرشاد	0	نشر ورجاوند و ایسته به شرکت ورجاوند	تاریخی خارجی	
یاسین نعیمی	4.50	یوسف علیخانی	0	نشر هرمند	ترساناک ایرانی	
یاسین نعیمی	4.60	پوستین گردر	1	نشر نیز	ترساناک خارجی	
یاسین طیب نعیمی	4.90	بوزف روت	1	نشر نون	چنایی و مهندسی ایرانی	
یاسین نعیمی	4.90	بو نیسو	1	نشر بو	چنایی و مهندسی خارجی	
وحید فقیهی	3.30	یعقوب بادعلی	0	نشر مهرگان خرد	روشنایی ایرانی	
وحید فقیهی	4.70	یاسمن حلیلی فرد	0	نشر مرکز	روشنایی خارجی	
وحید روشنی	4.20	پاروسلاو هاشک	1	نشر محتواهای الکترونیک صاد	طنز ایرانی	
وحید روحی	4.60	ویسنتون گروم	1	نشر هافن	طنز خارجی	
وحید آقاپور	4.40	ولنام فاکر	1	نشر ماه آوا	عائشگاه ایرانی	
هون خاطر	4.30	ولنام شنکیز	1	نشر کتاب هرمس	عائشگاه خارجی	
هونن شکیبا	4.40	ولنام سامرسن موام	1	نشر طرطره	علمی و تحقیقی ایرانی	
هونن شاطری پور، کامبیز خلیلی	4.80	ولنام پل یانگ، وین چیکویس، برد کامینکر	1	نشر شورآفرین	علمی و تحقیقی خارجی	
هونن شاطری پور، فربنا	4.30	ویکتور هرگو	1	نشر سنگ	فلسفی ایرانی	
قصیده‌نیاث، شهره روحی	4.30	ویرجینیا ووف	1	نشر روزگار نو	فلسفی خارجی	
هونن شاطری پور، فربنا	3.90	وودی آن	1	نشر روزگار	مذهبی ایرانی	
قصص‌نیاث، نایان		ونسا دینن باخ	1	نشر چنجه	مذهبی خارجی	

شکل ۵

داده‌ها در نرم افزار پاور بی آی مصورسازی شده‌اند. با استفاده از این نرم افزار می‌توان در ک بهتری از مصورسازی داده‌ها داشت و می‌توان در چند لحظه با استفاده از فیلترینگ از جهات مختلف داده‌ها را بررسی کرد. در تصویر بالا به ترتیب از چپ به راست، بهترین گوینده، بهترین نویسنده، انواع انتشاراتی که در دیتاست موجود هستند (ممکن است یک کتاب دارای چند انتشار مختلف باشد) و ژانرهای مختلف که همگی به دو دسته ایرانی و غیر ایرانی تقسیم شده‌اند به نمایش گذاشته شده است. در ادامه به تحلیل متغیرهای توضیحی مختلف این دیتاست پرداخته می‌شود تا بتوان به روابط بین این متغیرها بی‌برد.

۱.۲ تعداد آرا و امتیازات

از نظر رای دهنده‌گان این سری داده تصادفی سایت کتابراه، بهترین گویندگان که میانگین امتیاز آن‌ها به تعداد آرا آنها محاسبه شده است، گویندگانی هستند که در جدول نمایش داده شده‌اند. (آرش رستمی، آرمان سلطان‌زاده، ...)



Ketabrah.ir



Speaker	Ranking	Author	Translated	Publisher	Genre	Images
احسان چریکی، المیرا رحیمیان	5.00	پوناس پولانسن	1	وادجوان	اجتماعی ایرانی	
آرش رستمی	5.00	یوکیو میتئیما	1	لوین کتاب گویا	اجتماعی خارجی	
آرمان سلطان زاده	5.00	یوکو اوگاوا	1	نشر ویدا	تاریخی ایرانی	
اشکان عقیلی پور	5.00	پوسنگلی میرشکاک	0	نشر ورجاوند وابسته به شرکت ورجاوند	تاریخی خارجی	
اهنگ کلتزی	5.00	یوسف غلیدان	0	نشر طبله‌مند	ترسناک ایرانی	
امید تقی	5.00	یوسین گرد	1	نشر نهند	ترسناک خارجی	
امیر جلیلی	5.00	یوزف روت	1	نشر نون	جنایی و معتمد ایرانی	
بهنار بستان‌دوست	5.00	یعقوب بادعلی	0	نشر لو	جنایی و معتمد خارجی	
بلژن گنجعلی، شفیلا تائش، معین الدین شعاقی، امین عزرع زاده، صدق غرفی، امین ادروش	5.00	یاسمن خلیلی فرد	0	نشر مهرگان خرد	روانشناسی ایرانی	
مقدم، شهاب غزالی، افساله فتحی‌خواه، زهرا شیبانی، هادی مسعودی	5.00	پارسالاو هاشک	1	نشر مرکز	روانشناسی خارجی	
حافظ کاظم زاده	5.00	پولیام گووم	1	نشر محتواهای الکترونیک صاد	علمی و تحقیقی ایرانی	
حامد فقالی	5.00	پولیام سارمسرت موام	1	نشر ماهی	علمی و تحقیقی خارجی	
حسام الدین مقامی کیا	5.00	پولیام پل بانگ، وین چیکوپس، برد کامینگر	1	نشر ماه آوا	عاقلانه ایرانی	
حسن همایی	5.00	ونکتور هوگو	1	نشر هرمس	علمی و تحقیقی ایرانی	
حسین نصیری	5.00	ویرجینیا ولف	1	نشر قطره	علمی و تحقیقی خارجی	
رامونا شاه	5.00	وودی آلن	1	نشر شوراقدین	فلسفی ایرانی	
رامین پورامیان	5.00	ونسا دینیخ باخ	1	نشر سلگ	فلسفی خارجی	

شکل ۶

برای مثال از کتاب‌هایی که آرش رستمی در این دیتاست گوینده آن است می‌توان به کتاب و چشم‌هایش کهربایی بود انتشارات ققنوس اشاره کرد و یا از کتاب‌هایی که احسان چریکی و المیرا رحیمیان با هم در گویندگی آن همکاری کرده‌اند می‌توان به کتاب پیله‌ای چای بنوش انتشارات کتاب نیستان اشاره کرد.

به این اطلاعات می‌توان به سادگی در لینک قرار داده شده در پیوست (داشبورد مصورسازی داده در پاور) دسترسی پیدا کرد و با استفاده از فیلترهای تعییه شده در داشبورد به اطلاعات خاص هر نویسنده دسترسی پیدا کرد.



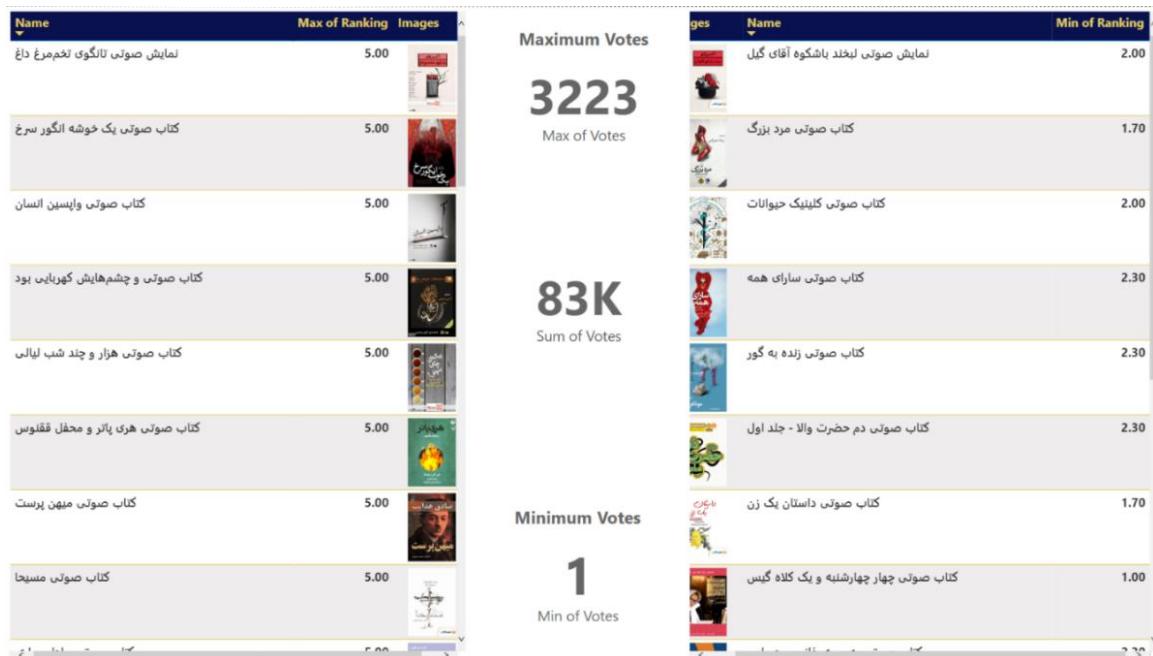
Ketabrah.ir



Speaker	Ranking	Author	Translated	Publisher	Genre	Images
بهاره رهنما	1.00	بهاره رهنما	0	نشر چشمه	داستان کوتاه ایرانی	
حسن آزادی	1.00					
رضا عمرانی، راضیه هاشمی	1.70					
شرکان اثنازاده	1.70					
رویا فلاخی	2.00					
علی آباقری	2.00					
گروه گویندگان	2.00					
آرمان سلطان زاده، غیریم پاک	2.30					
ذات						
آنالیز طاهریان	2.30					
رضا عزراخی، فربین منخدصی	2.30					
نورالدین چوادیان، نوشین حسن	2.30					
زاده						
بهاره قربانی پور	2.40					
نیما زینی	2.40					
هانیه محیی	2.40					
الهام کردا	2.50					
فریداد موسویان	2.50					
فریم حمذوب	2.50					
آزاده مویدی فرد	2.60					

شکل ۷

همچنین، کمترین امتیاز گویندگی به بهاره رهنما تعلق می‌گیرد که از کتاب‌های او به داستان چهارچهارشنبه و یک کلاه گیس می‌توان اشاره کرد.



شکل۱

همانطور که در تصویر بالا مشخص شده است، بیشترین تعداد آرا 3223^{13} و کمترین آن 1^{14} و در مجموع 13 هزار رای داده شده است.

بیشترین تعداد رای برای کتاب ملت عشق (اجرای جدید) از الیف شاکاف می‌باشد که گویندگی این کتاب توسط گروهی از گویندگان از جمله احمد گنجی، علی دنیوی ساروی، شیما درخشش، آرمان سلطان زاده، مجتبی مسافری نژاد، مریم پاک ذات، محسن زرآبادی پور، سعید مرزبالي، بهروز طالبي، مهدى قاسميان، مینو طوسی، پروین محمدیان، حامد حشمتی، میلاد فتوحی، میلاد میرزايو اجرا شده است که دارای ژانر عاشقانه و بیشترین نظرات را به خود جلب کرده است. در حالی که کمترین آرا برای چند کتاب مثل همه‌ی افق، واپسین انسان، یک شب دیگر هم و ... هست.

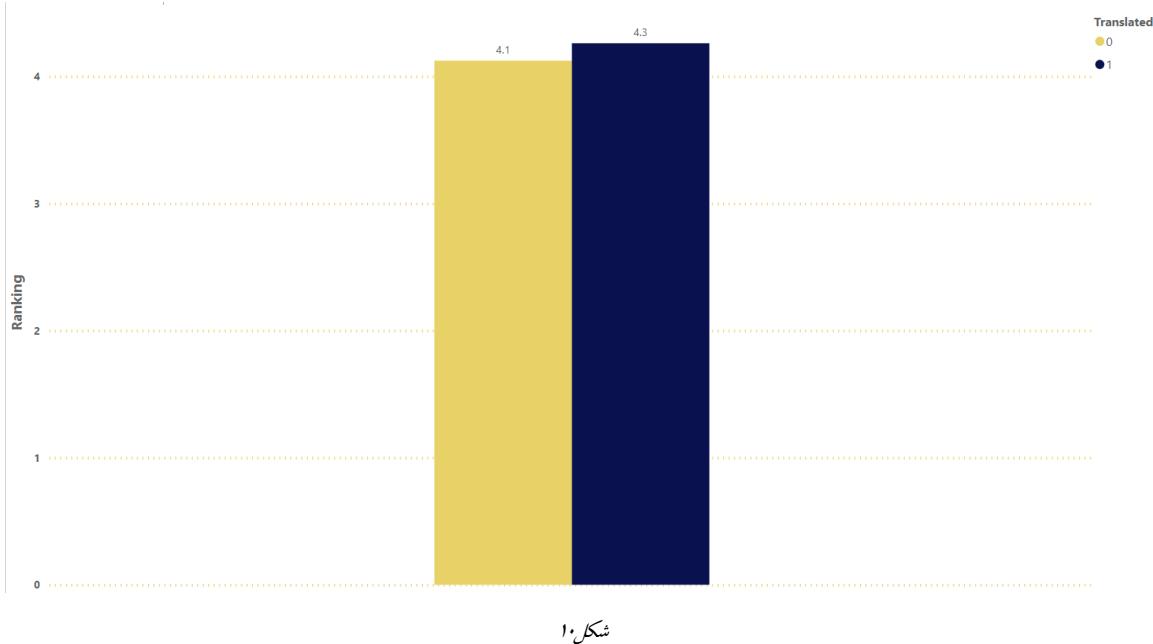


شکل۲

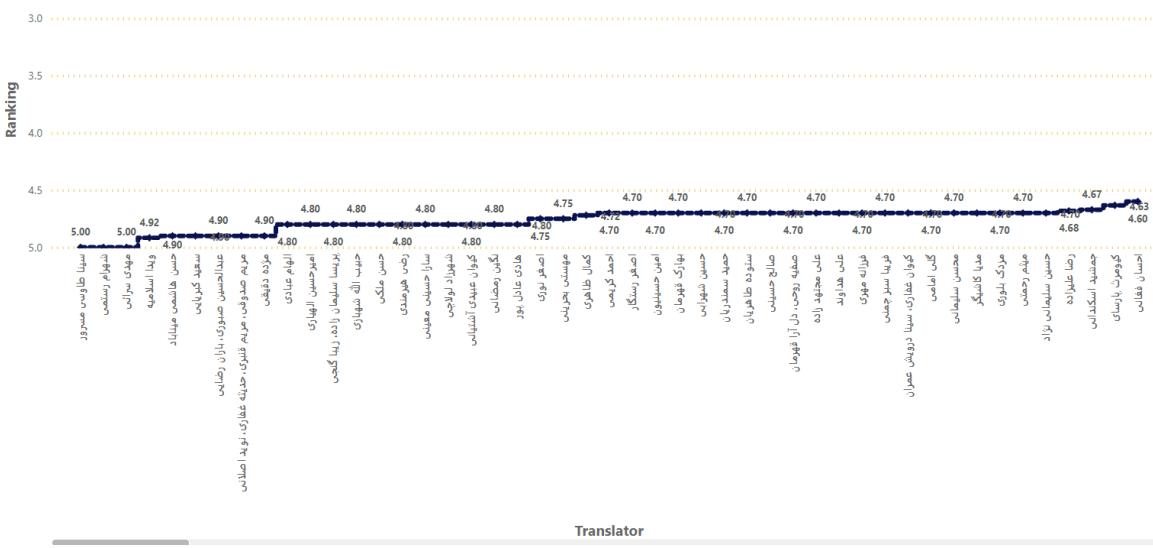
¹³ Maximum Votes

¹⁴ Minimum Votes

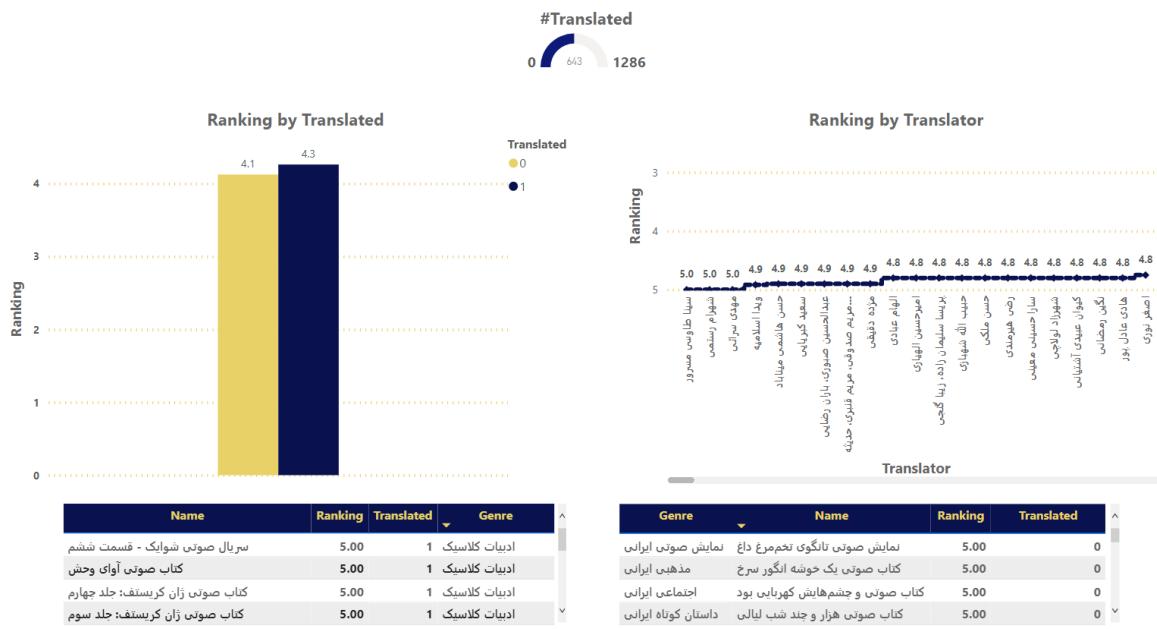
٢٠٢ ترجمہ



در کل باتوجه به نتایج بدست آمده، امتیاز کتاب‌های ترجمه شده (۲۰۳) بالاتر از امتیاز کتاب‌های ترجمه نشده (۱۶۰) می‌باشد و باتوجه به نظرات کاربران کتابهای نویسنده‌گان خارجی موفق‌تر از نویسنده‌گان ایرانی عمل کرده‌اند. اعداد مقایسه شده مجموع رنکینگ به تعداد کل آرا است. در شکل زیر اطلاعات محبوبیت مترجم‌های مختلف در دسترس می‌باشد.



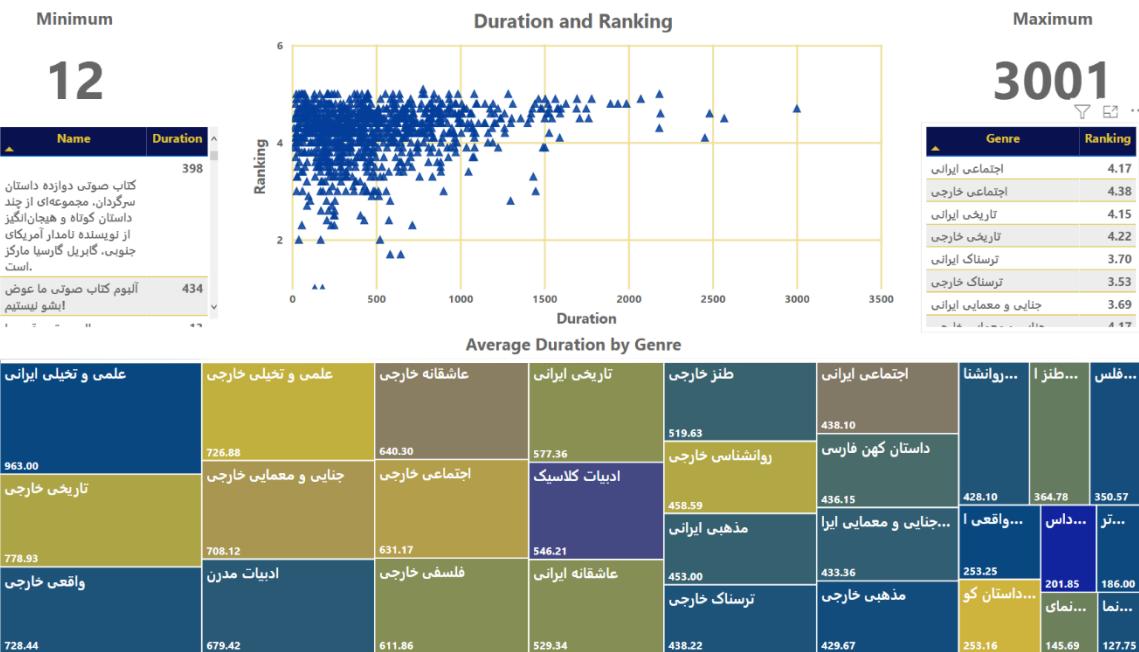
شکل ۱۱



شکل ۱۲

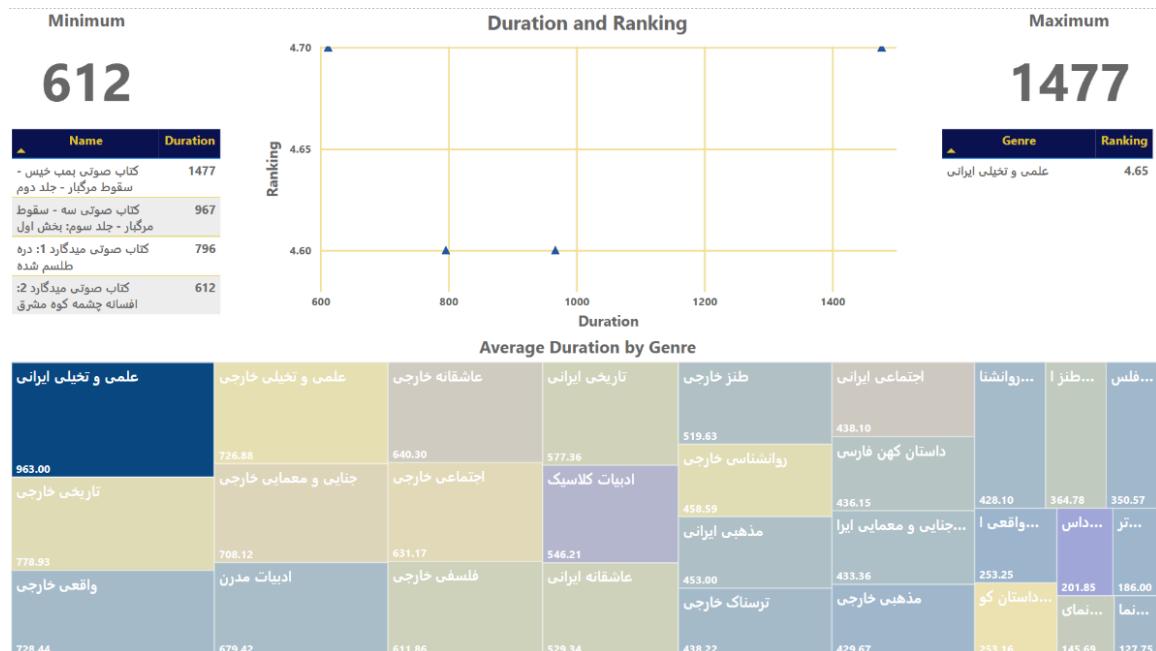
در مجموع ۶۴۳ کتاب ترجمه شده و ۳۴۰ کتاب ترجمه نشده در این دیتابست موجود می‌باشد. در کتاب‌های ترجمه شده، محبوب‌ترین کتاب‌ها، کتاب شوایک، آواز و حش، زان کریستف و ... می‌باشند و در کتاب‌های ترجمه نشده ایرانی کتاب‌های تانگوی تخم مرغ دار و یک خوش‌انگور سرخ و ... برترین کتاب‌ها در دو گروه ترجمه شده و نشده می‌باشند. رانر ادبیات کلاسیک در میان محبوب‌های ترجمه شده بسیار محبوب است.

۳.۲ مدت زمان

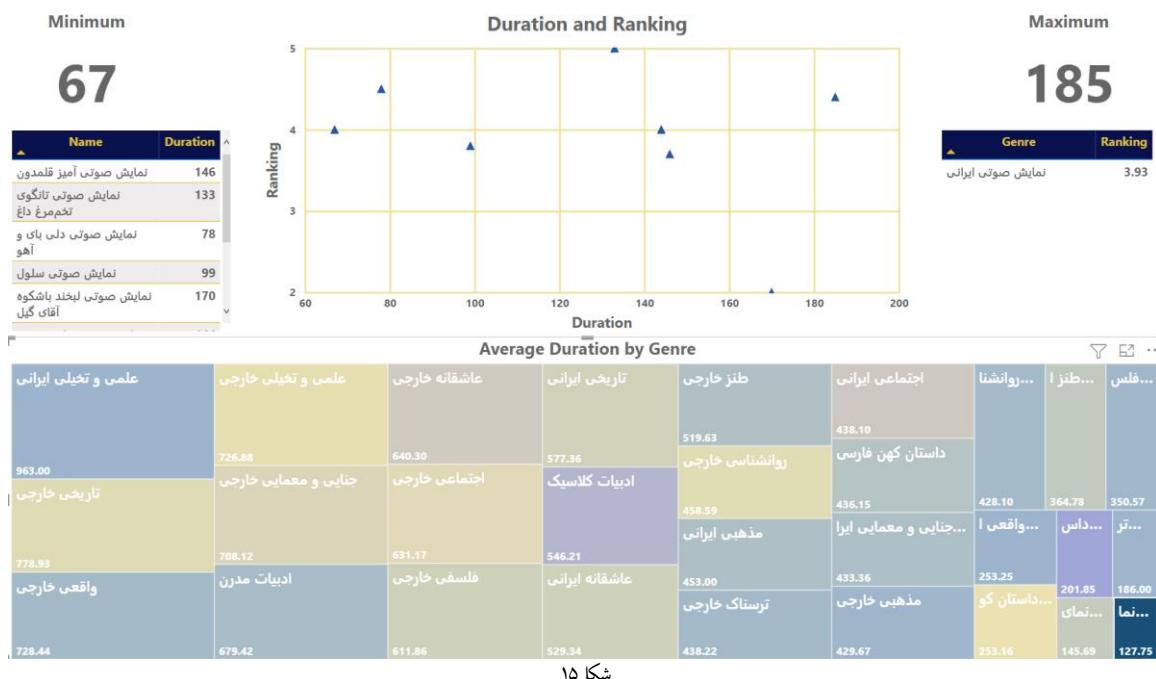


شکل ۱۳

بیشترین مدت زمان برای کتاب برادران کارامازوف (۳۰۰ دقیقه) که دارای ژانر فلسفی است و کمترین مدت زمان برای سریال صوتی رقص باعفتریت - قسمت دوم (۱۲ دقیقه) که دارای ژانر ترسناک است می‌باشد. با توجه به اسکرپلات^{۱۵} بالا، تقریباً نمی‌توان یک رابطه بین مدت زمان و امتیاز بالا پیدا کرد. هرچند در مدت زمان بالاتر از ۱۷۰۰ دقیقه امتیاز پایین تر از ۶۱۲ مشخص نشده است.



شکل ۱۴



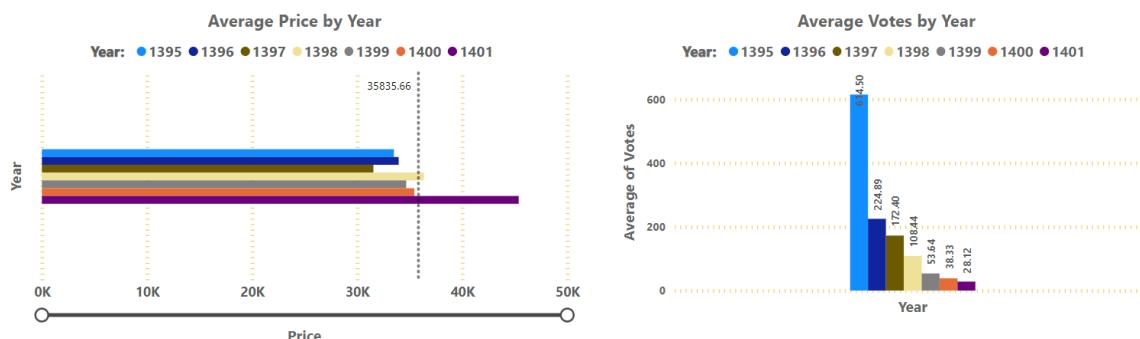
شکل ۱۵

¹⁵ Scatter Plot

میانگین بیشترین مدت زمان در نقشه درختی بالا^{۱۶} برای ژانر علمی و تخیلی ایرانی می‌باشد که در مجموع ۹۶۳ دقیقه می‌باشد که در این ژانر کتاب صوتی بمب خیس که نویسنده این کتاب معین فرد است بیشترین مدت زمان را به خود اختصاص داده است. تعداد آرا به این کتاب ۲۱۷ و امتیاز این کتاب ۲.۷ است. تعداد آرا برای این کتاب از میانگین آرا کمتر است.

همچنین ژانر نمایش صوتی ایرانی دارای میانگین کمترین مدت زمان برای ژانر علمی و تخیلی ایرانی می‌باشد. که در مجموع ۹۶۳ دقیقه می‌باشد که در این ژانر کتاب صوتی بمب خیس بیشترین مدت زمان را به خود اختصاص داده است. کتاب‌های دیگر در این ژانر نمایش صوتی آمیز قلمدون، نمایش صوتی تانگوی تخم مرغ دار، نمایش صوتی دلی بای و آهو و چند اثر دیگر هستند.

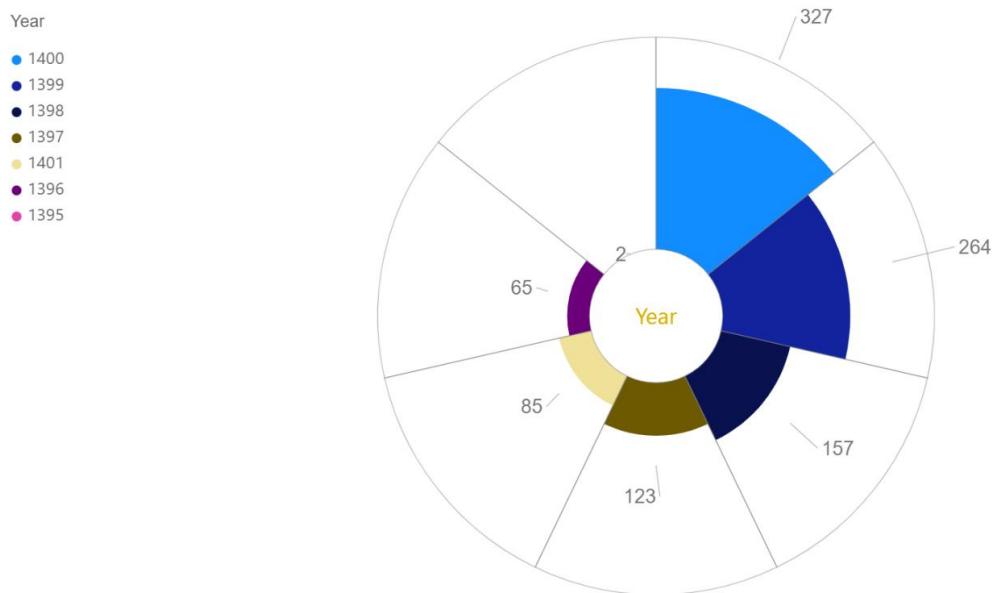
۴.۲ سال انتشار



شکل ۱۶

در سال ۱۳۹۵ کاربران بیشتر از بقیه سال‌ها به کتاب‌ها رای داده اند و میانگین قیمت کتاب‌ها تقریباً ۳۳۵۰۰ تومان بوده است. میانگین رای دهی در سال ۱۴۰۱ کمترین می‌باشد در حالی که میانگین قیمت در این سال رشد زیادی داشته است. کمترین میانگین قیمت نیز به سال ۱۳۹۷ اختصاص دارد که میانگین آرادر این سال حدود ۱۷۲.۵ می‌باشد. احتمال داده می‌شود با قیمت بالای کتاب، میل مخاطب به خواندن و خرید و در نهایت رای دادن به کتاب کم شده است.

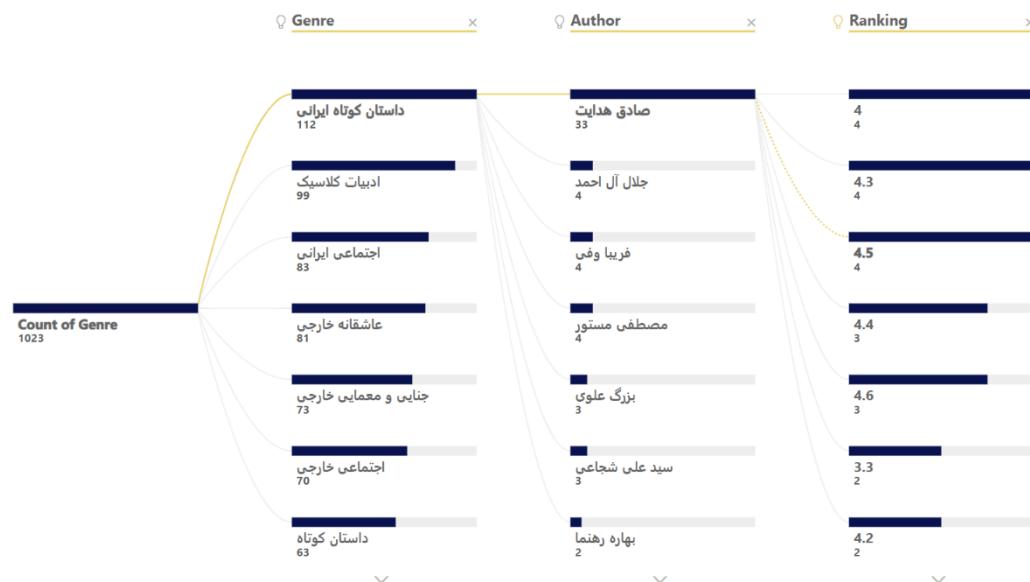
¹⁶ Tree Map



شکل ۱۷

باقطه به نمودار دونات بالا در سال‌های ۱۳۹۱ تا ۱۴۰۰ بصورت صعودی تعداد کتاب‌های منتشر شده زیاد شده است اما در سال ۱۴۰۱ افت داشته است.

۵.۲ نویسنده



شکل ۱۸

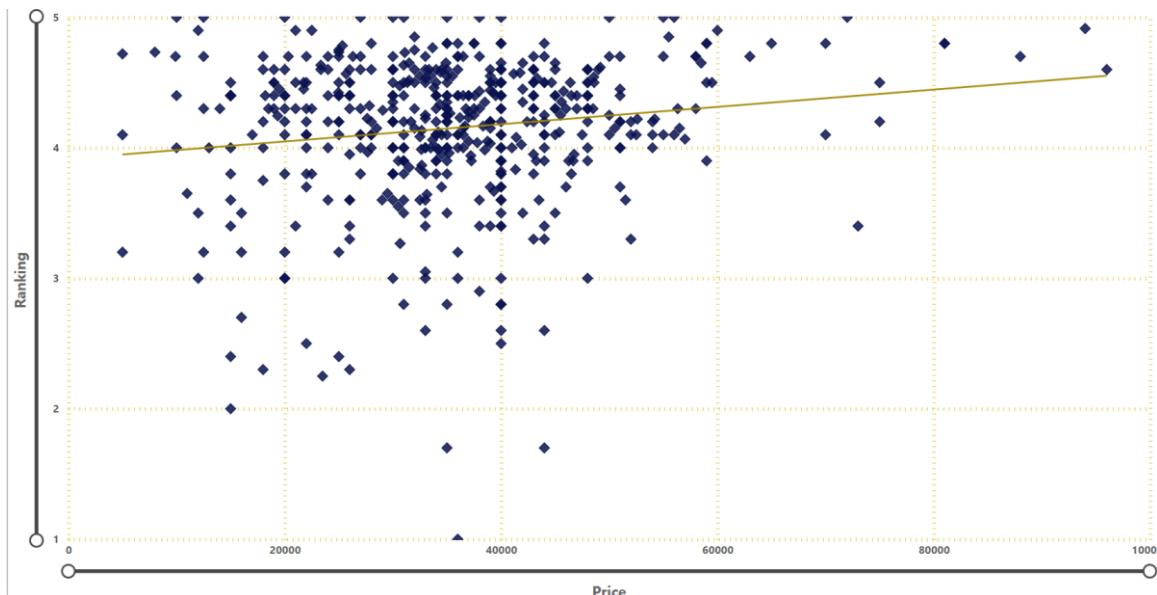
در این نمودار درختی^{۱۷} با انتخاب ژانر به نویسنده‌گانی که کتاب در این ژانر دارند دسترسی پیدا می‌کنیم و سپس با انتخاب نام نویسنده مثل صادق هدایت^{۳۲} کتاب در ژانر داستان کوتاه ایرانی) به رنکینگ آن نویسنده دسترسی پیدا می‌کنیم. در ستون اول فیلتر ژانر، فیلتر دوم فیلتر نویسنده و فیلتر ستون آخر مربوط به محبویت است. فیلترهای اول و دوم از بیشترین به کمترین مرتب شده‌اند. مثلاً بیشترین ژانر کتاب‌های ترجمه نشده به داستان کوتاه ایرانی تعلق دارد که صادق هدایت بیشترین کتاب را در این مورد دارد.

Most popular Author		Most popular Publisher		Most popular Speaker		Most popular Translator	
Author	Ranking	Publisher	Ranking	Ranking	Speaker	Ranking	Translator
احمد محمود	5.00	النشرات بخشایش	4.90	5.00	احسان جزیک، المیرا رحیمیان	5.00	سیننا طاووسی مسروور
آذرگفت بهرامی	5.00	النشرات شبک	4.90	5.00	ارش رستمی	5.00	شهرام رسنی
آندرzej سایکوفسکی	5.00	النشرات مولن	4.90	5.00	امیر جلیلی	5.00	مهدی سرانی
پاتریک دوبیت	5.00	النشرات شبک	4.85	5.00	بیهقی، شعبانی، عین الدین عشاقي،	4.92	ویدا اسلامیه
حسام الدین مقامی کیا	5.00	النشرات	4.85	برخ زاده صدف غریب، رادوش مقدم، ^{۱۸} افساله نلاحی، هزار شبانی، هادی	5.00	حسن هاشمی میناباد	4.90

شکل ۱۹

در این قسمت به محبوب‌ترین نویسنده، احمد محمود، محبوب‌ترین انتشارات، بخشایش، محبوب‌ترین گویندگان و محبوب‌ترین مترجم، سیننا طاووسی مسروور، شهرام رستمی، مهدی سرانی می‌باشد.

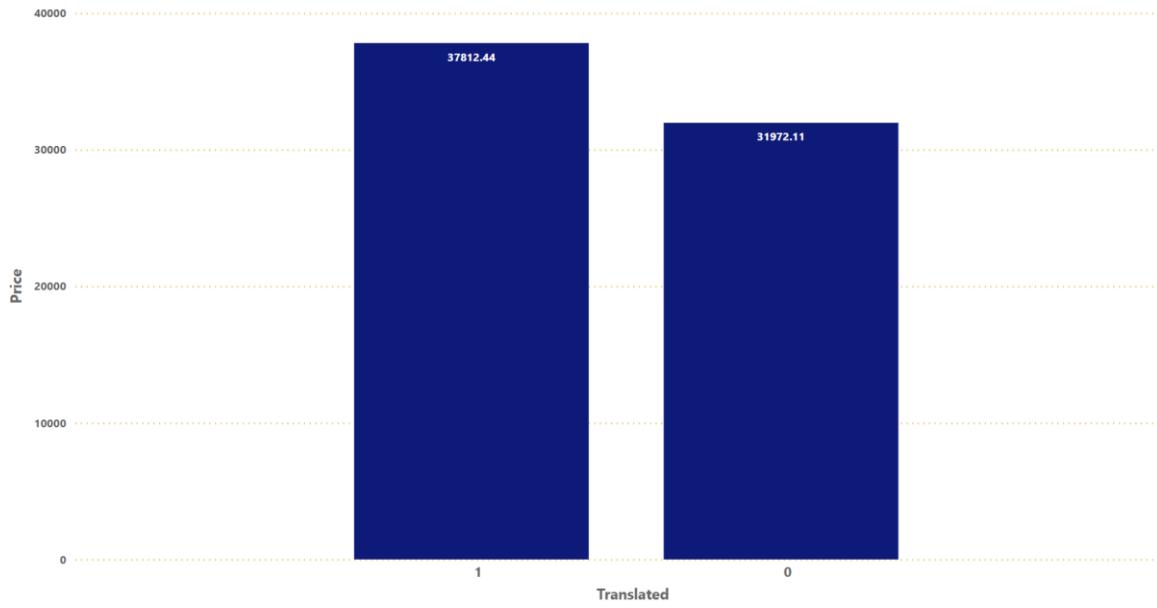
۶.۲ قیمت



شکل ۲۰

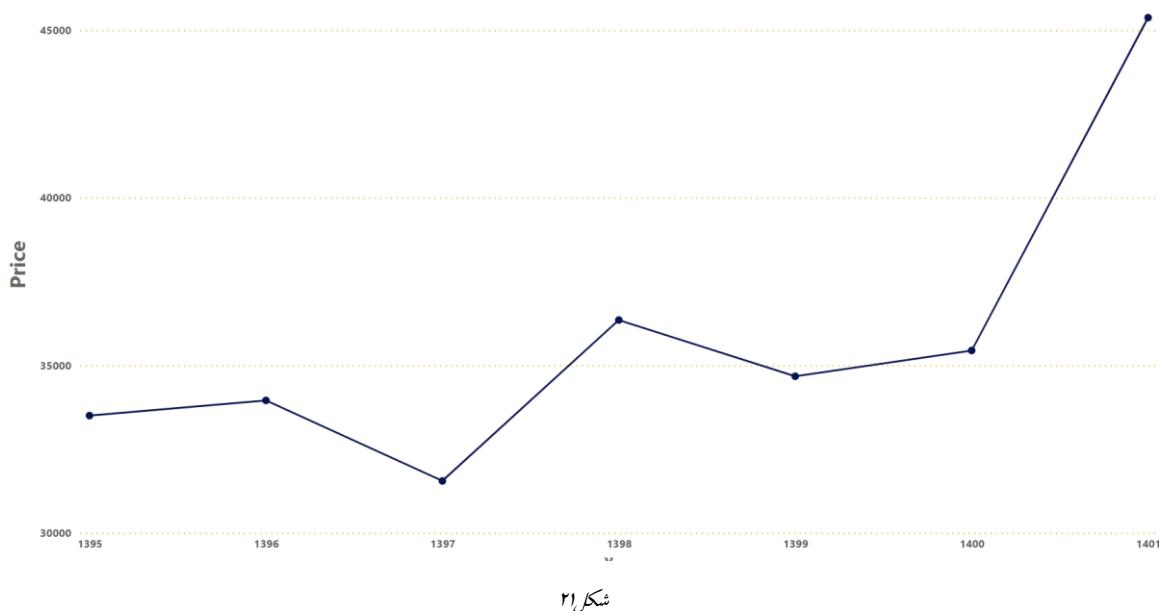
تقریباً بین قیمت کتاب‌ها و امتیاز آن‌ها یک رابطه صعودی می‌باشد که البته در این شکل می‌تواند تاثیر داده‌ی پرت هم باشد. هرچند در ناحیه قیمت ۳۵ تا ۴۰ هزار تومان امتیاز ۳۰ متراکم‌تر می‌باشد.

¹⁷ Decomposition Tree



شکل ۲۱

با توجه به هیستوگرام بالا و مقایسه قیمت کتاب‌های ترجمه شده و نشده متوجه می‌شویم قیمت کتاب‌های ترجمه شده (غیر ایرانی) بیشتر از کتاب‌های ایرانی می‌باشد.



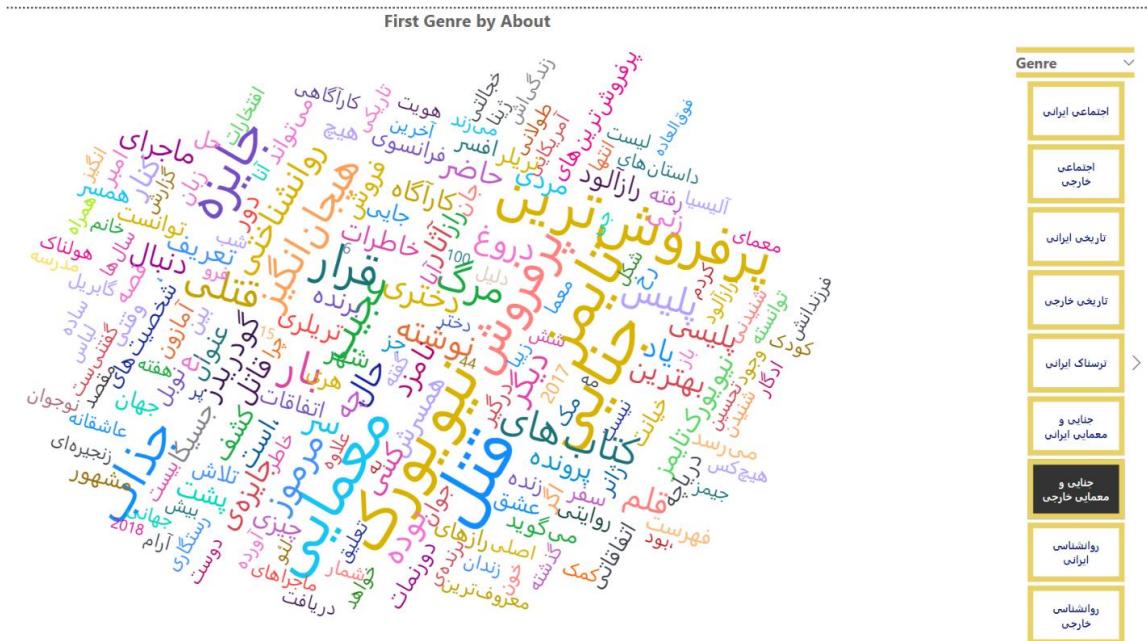
شکل ۲۱

در نمودار بالا می‌توان دریافت کرد در سال‌های متوالی قیمت کتاب در این سایت دچار نوساناتی شده است و از سال ۱۴۰۰ تا ۱۳۹۵ دچار افزایش زیادی شده است.

٧٢ خلاصہ کتاب



شکل ۲۳



۲۴

در نمودار ورد کلود^{۱۸} بالا بیشترین کلماتی که در توضیحات کتاب‌ها در ژانرهای مختلف آمده است نمایش داده شده است. در این قسمت کلمات اضافی^{۱۹} فارسی مانند از در به و ... حذف شده‌اند. اگر این کلمات را حذف نکنیم پر تکرار ترین کلمات هر ژانر کلماتی مانند از، تا، و ... می‌شوند که هیچ معنی خاصی ندارد. تصویر بالا ورد کلود ژانر تاریخی ایرانی است. در هر ژانر یک سری از کلمات کلاستر بندی شده‌اند که هرچه این کلمات بزرگ‌تر و پرنگ‌تر باشند دارای اهمیت بالاتری نیز هستند (یعنی در خلاصه‌ی کتاب‌های

¹⁸ Word Cloud

19 Stop Words

ژانر مربوطه پر تکرار ترین هستند). در لینک پاور بی آی گذاشته شده، می‌توان با انتخاب هر ژانر به این کلمات دسترسی پیدا کرد. در ژانر جنایی و معماهی پراهمیت‌ترین کلمات قتل، جنایی، معماهی، عجیب و ... هستند.

فصل سوم

پیش پردازش

در این بخش برخی از ویرایش‌های لازم را برای آماده شدن دیتا برای ورود داده به سیستم‌های پیشنهاد دهنده اعمال شده است. ابتدا نگاهی به داده‌ها در شکل اولیه خود بینداریم. داده‌ها در محیط ژوپیتر نوت بوک^{۲۰} قرار دارند. این دیتابست از ۱۵ ستون متفاوت تشکیل شده است. هر کدام از این ستون‌ها نمایشگر یک متغیر توضیحی جدا است و نوع هر متغیر باید در نهایت بصورت:

- نویسنده: کیفی، رسته‌ای^{۲۱}
- مترجم: کیفی، رسته‌ای
- گوینده: کیفی، رسته‌ای
- انتشارات: کیفی، رسته‌ای
- سال: کیفی، رسته‌ای
- فرمت: کیفی، حروف^{۲۲}
- زمان: کمی پیوسته، عدد^{۲۳}
- قیمت: کمی پیوسته، عدد
- ژانر: کیفی، رسته‌ای
- ترجمه: دودویی^{۲۴}
- عکس: کیفی، رسته‌ای
- خلاصه: کیفی، حروف
- نام کتاب: کیفی، رسته‌ای
- امتیازات: کمی پیوسته، عدد

می‌باشد.

Ranking	Name	About	Images	Votes	Translated	Genre	Price	Time	Format	Year	Publisher	Speaker	Translator	Author		
۲.۷	کتاب صوتی بزرگ و میش دانسته همچو	برگ و میش دانسته همچو	https://img.ketabrah.ir/img/8153564537814758...	1.0		(ز)ای کتابیک	۶.49	۲۱:۰۰ ساعت و دقیقه	۷ MP3	۱۳۹۷	متر دافت	احمد نگنچ	حسین عجمی	پتوتو کوبلو	۰	
۲.۸	کتاب صوتی لینیا چاه پریشانی است در بیوگرافی ...	لینیا چاه پریشانی است در بیوگرافی ...	https://img.ketabrah.ir/img/6701643057778123...	1.0		(ز)ای هزار	۹.99	۵۲:۰۰ ساعت و دقیقه	۱۹ MP3	۱۳۹۷	انتشارات قفس	احمد نگنچی، علی نبوی سریوش، سیما ترختی، ابوالشلاق	۱	
۲.۹	کتاب صوتی و تصویری هر انسانی، مانک مرگ - دوست و میوانان به درو	هر انسانی، مانک مرگ - دوست و میوانان به درو	https://img.ketabrah.ir/img/8558201548331221...	1.0		(ز)ای روانشناسی	۹.49	۵۱:۰۰ ساعت و دقیقه	۱۴ MP3	۱۳۹۷	نشر فرهاده	آرمان سلطان زاده	سیده حبیب	ارون بیرون	۲	
۲.۱۰	کتاب صوتی حضرت ما امام روحانی و پیغمبر میله داشتند و ...	حضرت ما امام روحانی و پیغمبر میله داشتند و ...	https://img.ketabrah.ir/img/8574475464041519...	1.0		(ز)ای هزار	۵.99	۷۸:۰۰ ساعت و دقیقه	۱۰ MP3	۱۴۰۰	انتشارات کتاب	علی و تحسین	فرهادی	محمد صالح وزیری زاده	مت هیگ	۳
۲.۱۱	کتاب صوتی رویانه گفت: تو هموز برای من سرمه‌ای بیش ...	رویانه گفت: تو هموز برای من سرمه‌ای بیش ...	https://img.ketabrah.ir/img/6790259418906252...	1.0		(ز)ای کتابیک	۱.89	۹:۰۰ ساعت و دقیقه	۲ MP3	۱۳۹۸	انتشارات پر	شهرزاد هسگری	مینم رحمتی	گروبریز	۴	

شکل ۲۵

²⁰ Jupyter Notebook

²¹ Categorical

²² String

²³ Integer

²⁴ One Hot Encoding

چند مورد ویرایش باید در داده‌ی اولیه اعمال شود:

- ✓ در ستون Time همه‌ی ساعت‌ها بهتر است به دقیقه تبدیل شود تا واحد یکتا داشته باشند چون در ادامه یک سری اعمال جبری روی این ستون باید اعمال شود:

Author	Translator	Speaker	Publisher	Year	Format	Time	Price	Genre	Votes	Translated	Images	About
0 پالتو کوبلیو	احمد گنجی	حسین نعیمی	نشر ثالث	۱۳۹۶	MP3	461	۷۴۰۰ ت - ۶.۴۹ بروز	ابنیات کلاسیک (زای) (۳۱۲۲ رای)	1.0	https://img.ketabrah.ir/img/l/8153564537814758...	پرسک «ستالین» نم داشت. هو گرگ و میان بود ...	
1 البف شافاک	امیر ارسلان	اصحیح	انتشارات فتوس	۱۳۹۷	MP3	1195	۵۴۰۰ ت - ۹.۹۹ بروز	عشناه خارجی (زای) (۳۲۲۳ رای)	1.0	https://img.ketabrah.ir/img/l/6701643057778123...	دنیا چاه پریمان است در بود نسیں. از پس رقش	
2 اروین پالوم	آرمن سیده هبیب	سلطان زاده	نشر فطره	۱۳۹۷	MP3	1009	۵۱۰۰ ت - ۹.۴۹ بروز	روانشناسی خارجی (زای) (۱۷۰۵ رای)	1.0	https://img.ketabrah.ir/img/l/8558201548331221...	هر - انسانی، مالک مرگ خود است و میتوان به ...	
3 مت هیگ محمدصلحت فاضله فراهانی	نور آنی زاده	انتشارات کتاب کوشه پیش	۱۴۰۰	MP3	628	۴۸۰۰ ت - ۵.۹۹ بروز	علمی و تطبیقی خارجی (زای) (۲۵۷ رای)	1.0	https://img.ketabrah.ir/img/l/8574475464041519...	حضرت‌ها ادامه‌ار بوند و پیشنهاد داشت و ...		
4 آزاد بارسفان	انتوان دو سنت اگزوپری	میثم رحمتی	شهرزاد سعکری	۱۳۹۸	MP3	121	۴۰۰ ت - ۱.۹۹ بروز	ابنیات کلاسیک (زای) (۶۵۷ رای)	1.0	https://img.ketabrah.ir/img/l/6790259418906252...	رویداد گفت: تو هیوز برای من پسرچهاری پیشنهاد پیش نیست...	
...
1134 آزاد بارسفان	دارد	میلان فتوخی	نشر چشم	۱۳۹۶	MP3	264	۵۱۰۰ ت - ۴.۰٪	اخلاصی ایرانی (زای) (۷ رای)	0.0	https://img.ketabrah.ir/img/l/2368758634487696...	این دادستان در منت زمان تو سال روایت میشود و ...	

شکل ۱۶

- ✓ همه‌ی قیمت‌ها در ستون Price باید به تومان تبدیل شوند تا مقیاس واحد داشته باشند:

Author	Translator	Speaker	Publisher	Year	Format	Time	Price	Genre	Votes	Translated	Images	About
0 پالتو کوبلیو	احمد گنجی	حسین نعیمی	نشر ثالث	۱۳۹۶	MP3	461	۷۴۰۰ ت	ابنیات کلاسیک (زای) (۳۱۲۲ رای)	1.0	https://img.ketabrah.ir/img/l/8153564537814758...	پرسک «ستالین» نم داشت. هو گرگ و میان بود ...	
1 البف شافاک	امیر ارسلان	اصحیح	انتشارات فتوس	۱۳۹۷	MP3	1195	۵۴۰۰ ت	عشناه خارجی (زای) (۳۲۲۳ رای)	1.0	https://img.ketabrah.ir/img/l/6701643057778123...	دنیا چاه پریمان است در بود نسیں. از پس رقش	
2 اروین پالوم	آرمن سیده هبیب	سلطان زاده	نشر فطره	۱۳۹۷	MP3	1009	۵۱۰۰ ت	روانشناسی خارجی (زای) (۱۷۰۵ رای)	1.0	https://img.ketabrah.ir/img/l/8558201548331221...	هر - انسانی، مالک مرگ خود است و میتوان به ...	
3 مت هیگ محمدصلحت فاضله فراهانی	نور آنی زاده	انتشارات کتاب کوشه پیش	۱۴۰۰	MP3	628	۴۸۰۰ ت	علمی و تطبیقی خارجی (زای) (۲۵۷ رای)	1.0	https://img.ketabrah.ir/img/l/8574475464041519...	حضرت‌ها ادامه‌ار بوند و پیشنهاد داشت و ...		
4 آزاد بارسفان	انتوان دو سنت اگزوپری	میثم رحمتی	شهرزاد سعکری	۱۳۹۸	MP3	121	۴۰۰ ت	ابنیات کلاسیک (زای) (۶۵۷ رای)	1.0	https://img.ketabrah.ir/img/l/6790259418906252...	رویداد گفت: تو هیوز برای من پسرچهاری پیشنهاد پیش نیست...	
...
1134 آزاد بارسفان	دارد	میلان فتوخی	نشر چشم	۱۳۹۶	MP3	264	۵۱۰۰ ت	اخلاصی ایرانی (زای) (۷ رای)	0.0	https://img.ketabrah.ir/img/l/2368758634487696...	این دادستان در منت زمان تو سال روایت میشود و ...	
1135 فرهنگ آفاین	مریم نوری	دارد	درهشان فتوس	۱۳۹۷	MP3	656	۵۱۰۰ ت	اخلاصی ایرانی (زای) (۷ رای)	0.0	https://img.ketabrah.ir/img/l/3381286592472744...	کتاب صوتی از شیطان اموهه و سوران	

شکل ۱۷

✓ در ستون Votes، آرای باید به اعداد تبدیل شوند:

Author	Translator	Speaker	Publisher	Year	Format	Time	Price	Genre	Votes	Translated	Images	About
0	پالولو کوتولو	احمد گنجی	حسن نعیم	۱۳۹۶	نشر نات	MP3	461	۴۴۰۰	اندیشه	3142	1.0 https://img.ketabrah.ir/img/l/8153564537814758...	پرسک «سندنگو» نم داشت هر گزگز و میش بود که...
1	لطف شفاق	امحمد گنجی، علی نبیوی، ساروی، نیما در هشت، فتوس	ارسان فصیحی	۱۳۹۷	نشر نات	MP3	1195	۵۴۰۰	عشقانه	3223	1.0 https://img.ketabrah.ir/img/l/6701643057778123...	دلخواه بریشید است بر بو شدم. از پس رقش ...
2	ارون اژدها	آرمان سیده حبیب	سلطان زاده	۱۳۹۷	نشر نظره	MP3	1009	۵۱۰۰	روانشناسی	1705	1.0 https://img.ketabrah.ir/img/l/8558201548331221...	فر - دانشمند، مالک مرگ خود است به میتواند به روز ...
3	مت هیگ	فاضمه فراهانی	محمد صالح نورانی زاده	۱۴۰۰	انتشارات کوله پیش	MP3	628	۲۸۰۰	علمی و تحقیقی	657	1.0 https://img.ketabrah.ir/img/l/8574475464041519...	حریت‌ها ادامه‌وار بودند پیشنهاد دانسته و ...
4	آنوار نویسنده	آرمان شهزاده	مینم رحمنی اکبرپوری	۱۳۹۸	انتشارات پر	MP3	121	۹۰۰	اندیشه	957	1.0 https://img.ketabrah.ir/img/l/6790259418906252...	رویداد گفت: تو هیز برای من پسرچهاری پیش لیست ...
...
1134	آرمان بارسین	سازاره	میلانه فتوحی	۱۳۹۶	نشر چشم	MP3	264	۳۱۰۰	اخلاصی	7	0.0 https://img.ketabrah.ir/img/l/2368758634487696...	کتاب صوتی از شیخان اموهات و سواند ...
1135	فرخنده اذیان	نادر	مریم نوری درخشن	۱۳۹۷	انتشارات فتوس	MP3	656	۴۴۰۰	اخلاصی	7	0.0 https://img.ketabrah.ir/img/l/3381286592472744...	...

شکل ۲۸

در نهایت، دیتاست مرتب شده به شکل زیر قابل استفاده می‌باشد. این دیتاست چون بصورت دستی استخراج شده است، داده گمشده ۲۵ ندارد.

Author	Translator	Speaker	Publisher	Year	Format	Time	Price	Genre	Votes	Translated	Images	About	Name	Ranking
0	پالولو کوتولو	احمد گنجی	حسن نعیم	۱۳۹۶	نشر نات	MP3	461	۳۴۰۰۰	اندیشه	3142	1.0 https://img.ketabrah.ir/img/l/8153564537814758...	پرسک «سندنگو» نم داشت هر گزگز و میش بود که...	کیمیک	4.7
1	لطف شفاق	امحمد گنجی، علی نبیوی، ساروی، نیما در هشت، فتوس	ارسان فصیحی	۱۳۹۷	نشر نات	MP3	1195	۵۴۰۰۰	عشقانه	3223	1.0 https://img.ketabrah.ir/img/l/6701643057778123...	دلخواه بریشید است بر بو شدم. از پس رقش ...	ملت عشق (اجرا چیند)	4.8
2	ارون اژدها	آرمان سیده حبیب	سلطان زاده	۱۳۹۷	نشر چشم	MP3	1009	۵۱۰۰۰	روانشناسی	1705	1.0 https://img.ketabrah.ir/img/l/8558201548331221...	فر - دانشمند، مالک مرگ خود است به میتواند به روز ...	گرفت خواهش میتواند روز ...	4.8
3	مت هیگ	فاضمه فراهانی	محمد صالح نورانی زاده	۱۴۰۰	انتشارات کوله پیش	MP3	628	۲۸۰۰۰	علمی و تحقیقی	657	1.0 https://img.ketabrah.ir/img/l/8574475464041519...	حریت‌ها ادامه‌وار بودند پیشنهاد دانسته و ...	کتابخانه نیمه ...	4.6
4	آنوار نویسنده	آرمان شهزاده	مینم رحمنی اکبرپوری	۱۳۹۸	انتشارات پر	MP3	121	۹۰۰۰	اندیشه	957	1.0 https://img.ketabrah.ir/img/l/6790259418906252...	رویداد گفت: تو هیز برای من پسرچهاری پیش لیست ...	شزاده کوچولو من پسرچهاری پیش لیست	4.7
...
1134	آرمان بارسین	سازاره	میلانه فتوحی	۱۳۹۶	نشر چشم	MP3	264	۳۱۰۰۰	اخلاصی	7	0.0 https://img.ketabrah.ir/img/l/2368758634487696...	این دادستان نه مخت... مان نو رسان و پیش میتواند و ...	پیشنهاد	3.9
1135	فرخنده اذیان	نادر	مریم نوری درخشن	۱۳۹۷	انتشارات فتوس	MP3	656	۴۴۰۰۰	اخلاصی	7	0.0 https://img.ketabrah.ir/img/l/3381286592472744...	این دادستان نه مخت... مان نو رسان و پیش میتواند و ...	از شیطان اموهات و سوزاده فرخنده	4.7
1136	نهضه شاه	میلانه فتوحی	نهضه سوزده	۱۳۹۷	پهلوان سانتوس	MP3	432	۱۲۰۰۰	اخلاصی	7	0.0 https://img.ketabrah.ir/img/l/5205101502856184...	کتاب صوتی پویاندر لر نهضه شاه میخواهد... نهضان	نویسندر ...	4.9
1137	صلیلی	میرم مدرس	میرم مدرس فاطمی	۱۳۹۹	نشر مركز	MP3	494	۴۴۰۰۰	اخلاصی	6	0.0 https://img.ketabrah.ir/img/l/6683276132142126...	کتاب صوتی خداحضر ازینهش نویسنده خداحضر مدرس صلیلی	هدفهای ارزیشی ...	4.1
1138	نهضه شاه	علی شاه	محسن خدای	۱۳۹۷	انتشارات سوزده	MP3	822	۳۰۰۰۰	اخلاصی	1	0.0 https://img.ketabrah.ir/img/l/3996689476700875...	در کتاب صوتی خداحضر شاهلگان توشانی هی شاهله نه	دار شاهلگان	4.5

1023 rows x 15 columns

شکل ۲۹

۲۵ Missing Value

```

Any null values in Books:
Author      0
Translator   0
Speaker     0
Publisher   0
Year        0
Format      0
Time        0
Price       0
Genre        0
Votes       0
Translated  0
Images      0
About       0
Name        0
Ranking     0
dtype: int64
Books Shape: (1023, 15)
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1023 entries, 0 to 1138
Data columns (total 15 columns):
 #   Column    Non-Null Count Dtype  
---- 
 0   Author     1023 non-null   object  
 1   Translator 1023 non-null   object  
 2   Speaker    1023 non-null   object  
 3   Publisher  1023 non-null   object  
 4   Year       1023 non-null   category 
 5   Format     1023 non-null   object  
 6   Time       1023 non-null   int64  
 7   Price      1023 non-null   object  
 8   Genre      1023 non-null   object  
 9   Votes      1023 non-null   object  
 10  Translated 1023 non-null   float64 
 11  Images     1023 non-null   object  
 12  About      1023 non-null   object  
 13  Name       1023 non-null   object  
 14  Ranking    1023 non-null   object  
dtypes: category(1), float64(1), int64(1), object(12)
memoryv usage: 121.2+ KB

```

شکل ۳.

در مورد دیتاست نهایی یک سری اطلاعات بدست آورده شده است. نوع هر متغیر در خروجی بالا آورده شده است. برخی از این متغیرها باید بصورت رمزگذاری-تک داغ^{۲۶} [۳] وارد دیتاست شوند و باید نوع آنها را به رسته‌ای تغییر داد. در یادگیری ماشینی و سیستم‌های توصیه گر، رمزگذاری تک داغ روشی است که اغلب برای مقابله با داده‌های طبقه‌بندی (رسته‌ای) استفاده می‌شود. از آنجایی که بسیاری از مدل‌های یادگیری ماشینی نیاز دارند که متغیرهای ورودی آن‌ها عددی باشد، متغیرهای طبقه‌بندی باید در قسمت پیش‌پردازش تبدیل شوند. داده‌های طبقه‌بندی می‌توانند اسمی یا ترتیبی باشند. داده‌های ترتیبی دارای ترتیب رتبه‌ای برای مقادیر خود هستند و بنابراین می‌توانند از طریق رمزگذاری ترتیبی به داده‌های عددی تبدیل شوند. نمونه‌ای از داده‌های ترتیبی، رتبه‌بندی‌های آزمونی از A تا F است که می‌تواند با استفاده از اعداد از ۱ تا ۶ رتبه‌بندی شود. از آنجایی که هیچ رابطه کمی بین مقادیر فردی متغیرهای اسمی وجود ندارد، استفاده از رمزگذاری ترتیبی به طور بالقوه می‌تواند یک داستان خیالی ایجاد کند. رابطه ترتیبی در داده‌ها. بنابراین، رمزگذاری تک داغ اغلب برای متغیرهای اسمی به منظور بهبود عملکرد الگوریتم اعمال می‌شود. برای هر مقدار منحصر به فرد در ستون دسته بندی اصلی، یک ستون جدید در این روش ایجاد می‌شود. سپس این متغیرهای ساختگی با صفر و یک پر می‌شوند (۰ به معنای FALSE، ۱ به معنای TRUE).

One Hot Encoding

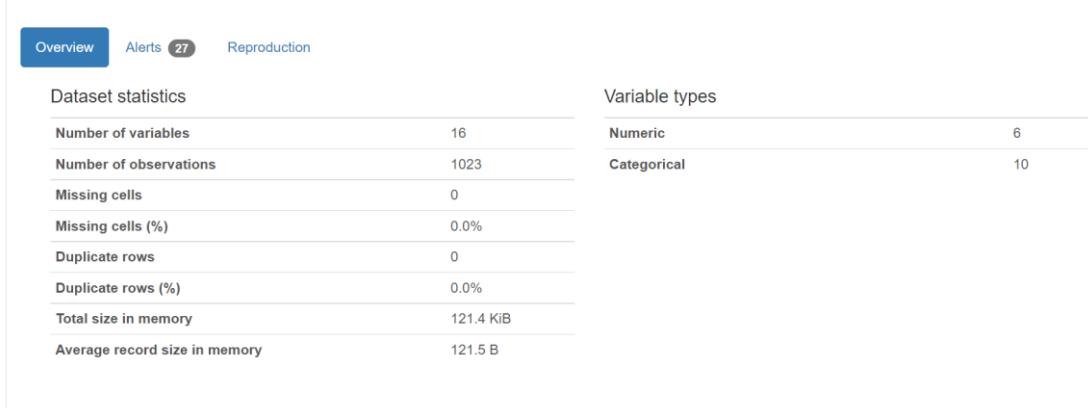
Apple	Chicken	Broccoli	Calories
1	0	0	95
0	1	0	231
0	0	1	50

شکل ۴.

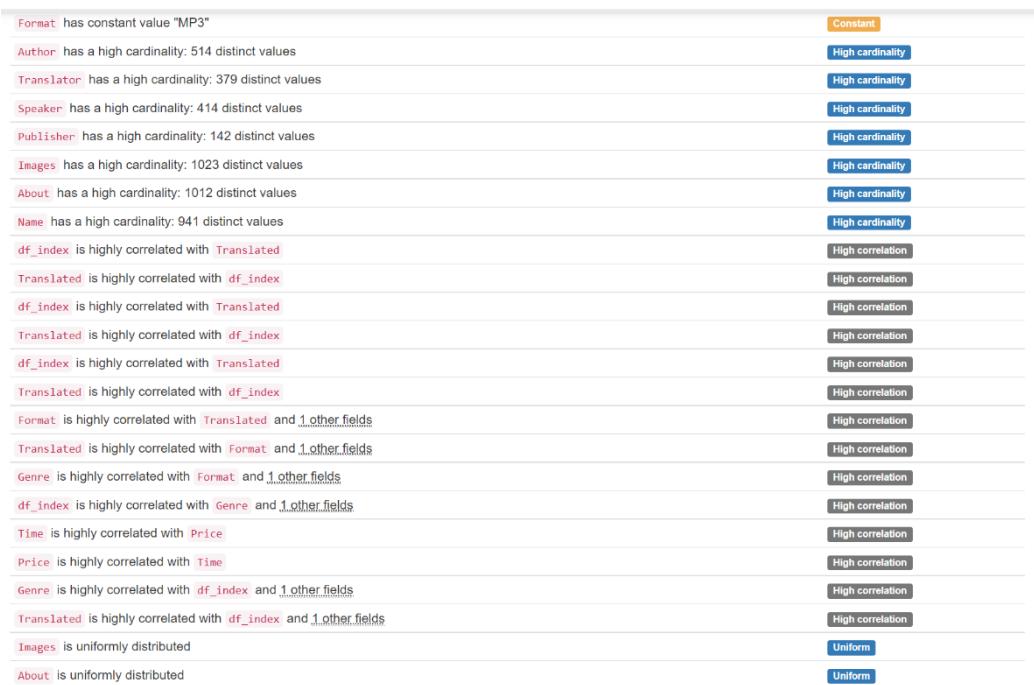
²⁶ One-Hot Encoding

بخش دیتا پروفایلینگ^{۲۷} نیز اطلاعات جزئی در مورد داده در اختیار قرار داده است.

Overview



شکل ۳۲



شکل ۳۳

چند نکته در کاردینالیتی^{۲۷} بالا ذکر شده است. مانند یکسان بودن مقادیر ستون فرمت^{۲۸}. همچنین، متغیرهای کیفی مانند نویسنده و مترجم و گوینده که دارای تعداد متوجهی مقدار منحصر به فرد هستند، می‌توانند پیچیدگی مدل را بالا ببرند. به علاوه، متغیرهایی مانند زمان و قیمت نیز می‌توانند همبستگی بالا داشته باشند.

²⁷ Cardinality

²⁸ Format

فصل چهارم

سیستم پیشنهاد دهنده کتاب

ریکامندر سیستم [4] به پیشنهاد محصولات، خدمات، محتوا یا دیگر اقلام به کاربر، بر اساس علائق او می‌پردازد. ریکامندر سیستم صرفاً اقلامی را به طور خود کار به کاربر پیشنهاد می‌کند که کاربر ممکن است آنها را پسندد. ریکامندر سیستم ها در حال حاضر بسیار متداول هستند و بسیاری از افراد از آنها استفاده می‌کنند بدون آنکه بدانند آنچه به آنها کمک می‌کند تا کالا، خدمات و محتوای دلخواه خود را سریع‌تر پیدا کنند، یک ریکامندر سیستم است. با توجه به آنکه کاربران احتمالاً نمی‌توانند همه محصولات یا محتوای موجود در یک وب‌سایت را مشاهده کنند، ریکامندر سیستم نقش مهمی را در کمک به آنها برای داشتن تجربه بهتر ایفا می‌کند و در عین حال، محتوایی را در اختیار کاربر قرار می‌دهد که شاید کاربر در فرایند اکشافی خود هرگز به آنها دست پیدا نکند. همچون بسیاری از موضوعات و مباحث علمی، در بحث سیستم پیشنهاد دهنده هوشمند نیز دانشمندان گوناگون تعاریف متعددی را برای ریکامندر سیستم ارائه کرده‌اند. در اینجا، برخی از تعاریف ارائه شده در میان چیستی ریکامندر سیستم با هدف بیان مفهوم آن به طور دقیق، تشریح شده‌اند.

سیستم‌های پیشنهاد دهنده بر اساس سه رویکرد اصلی زیر ساخته می‌شوند:

- روش‌های مشارکت محور²⁹ یا پالایش گروهی
- روش‌های محتوا محور³⁰
- روش‌های ترکیبی: در این روش از ترکیب روش‌های مشارکت محور و محتوا محور استفاده می‌شود.

انواع سیستم پیشنهاد دهنده مبتنی بر الگوریتم‌های مشارکت محور یا پالایش گروهی بر اساس تعاملات گذشته کاربران و انتخاب‌های پیشین آنها، توصیه‌های جدید ارائه می‌دهند. این تعاملات در ساختاری با عنوان «ماتریس تعاملات کاربر-آیتم» ذخیره می‌شوند. ایده اصلی حاکم بر روش‌های مشارکت محور این است که تعاملات گذشته کاربر-آیتم برای شناسایی موارد مشابه و پیشنهادهای بعدی کفایت می‌کند. پالایش مشارکتی یکی از روش‌های مطرح در این زمینه است که با تحلیل امتیازات کاربران سعی می‌کند امتیاز بعدی را بر اساس کاربران مشابه پیش‌بینی کند. الگوریتم‌های مشارکت محور به دو زیر مجموعه اصلی زیر تقسیم می‌شوند. در این پژوهه چون در مجموعه داده دسترسی به گذشته کاربران این سایت و انتخاباتشان نداریم از روش محتوا محور استفاده می‌شود.

²⁹ Collaborative Filtering

³⁰ Content-Based Methods

در ابتدا داده خود را برای این روش آماده می کنیم:

Author	Translator	Speaker	Publisher	Year	Format	Time	Price	Genre	Votes	Translated	Images	About	Name	Ranking
0 پالتو کوشاو	حسین نعیس	امحمد گنجی	نشر تلت	1396	MP3	461	34000	Classical literature	3142	1	https://img.ketabrah.ir/img/8153564537814758...	The boy's name was "Santiago". It was dusk when he...	مسکن	4.7
1 ارسطون فرموده	الیف شاپلک	امحمد گنجی، علی نبوی	ارشاد انتشارات قفس	1397	MP3	1195	54000	Romantic	3223	1	https://img.ketabrah.ir/img/6701643057778123...	The world is a well of distress in the absence...	مل عشق	4.8
2 اروین بایلم	سیدهدیه حبیبی	آرمان سلطان زاده	نشر فطره	1397	MP3	1009	51000	Psychology	1705	1	https://img.ketabrah.ir/img/8558201548331221...	- Every human being is the owner of his own de...	وقتی بدهی گرفتست	4.8
3 مه هیگ	محمد صالح	نور آن راه	فاطمه فراهانی	1400	MP3	628	28000	Science fiction	657	1	https://img.ketabrah.ir/img/8574475464041519...	The longings were continuous and foregrounded ...	کانگنهای سینه شب	4.6
4 انتون پوچت	کلارکی	میرم رحمنی	شهرزاد منگزی	1398	MP3	121	9000	Classical literature	957	1	https://img.ketabrah.ir/img/6790259418900252...	*The fox said: You are still nothing more than...	شارل کوکو	4.7

٣٤

در این قسمت، دیتابست موجود باید دچار تغییراتی شود تا بتوان آن را وارد سیستم توصیه گر کرد. ستون *About* و *Genre* را با کتابخانه گوگل ترنسلیتر^{۳۱} به زبان انگلیسی باز گردانده است تا ساده‌تر و دقیق‌تر بتوان از تکیک‌های مرتبط با پردازش متن^{۳۲} استفاده کرد.

روند ساخت سیستم پیشنهاد دهنده:

۱-متغیرهایی که می‌خواهیم در معیار شاخص تاثیرگذار باشند را جدا می‌کنیم. متغیرهای ژانر، خلاصه داستان، گوینده و نویسنده متغیرهایی هستند که یک کاربر اگر کتابی را پسندید مایل است کتاب بعدی با این ویژگی‌ها مشابه و یا نزدیک این ویژگی‌ها باشد.

```
In [116]: M booksen['tags'] = booksen['Genre']+ ' '+booksen['About']+ ' '+booksen["Speaker"]+ ' '+ booksen['Author']+ ' '+ booksen['Genre']+ ' '+booksen['Genre']

In [117]: M booksen.tags[0]

Out[117]: 'Classical literature The boy's name was "Santiago". It was dusk when he arrived at an abandoned church with his sheep. The roof of the church collapsed and a large plantain tree had grown in the place that was previously used for sacred vessels.\nHe decided to spend the night there. He brought all his sheep back inside the church and blocked the door with some poles and boards so that no sheep could leave during the night. There was no trace of a wolf or a bear that night when one of his sheep was separated from the flock, he had to spend the whole next day looking for that animal. He swept the floor with his hands and lacing the book he had finished a while ago on his pillow. "I must start reading thicker books," she said to herself, "because they will both take more comfortable pillows." سانشیو کی کتاب کا ملکہ کے پروردگار کا کہانی
```

٣٥ شکل

۲- از تکنیک ریشه یابی^۳ برای تبدیل کلمات مشابه به یکدیگر برای بهتر شناسایی کردن کلمه استفاده می‌کنیم. این مرحله تکنیک نرم‌السازی زبان شناسی است و برای تبدیل لغت به فرم ریشه بکار می‌رود. مثلاً لغت *Arrived* به لغت *Arriv* تبدیل می‌شود. ریشه پایی حسبت؟

ریشه در پردازش زبان طبیعی (NLP)، به فرآیند کاهش یک کلمه به ریشه کلمه آن اشاره دارد که به پسوندها و پیشوندها یا ریشه‌ها اضافه می‌شود. در حالی که یک الگوریتم ریشه یک فرآیند عادی سازی زبانی است که در آن اشکال مختلف یک کلمه به شکل استاندارد کاهش می‌یابد. این تکنیکی است که برای استخراج شکل پایه کلمات با حذف پسوندها از آنها استفاده می‌شود. درست مثل بریدن شاخه‌های یک درخت تا ساقه اش. به عنوان مثال، ریشه کلمات «خوردن»، «خورده»، «خوردن» است. موتورهای جستجو از stemming برای نمایه سازی کلمات استفاده می‌کنند. به همین دلیل است که به جای ذخیره کردن همه اشکال یک کلمه، یک موتور جستجو می‌تواند فقط ریشه‌ها را ذخیره کند. به این ترتیب ساقه زنی باعث کاهش اندازه شاخص و افزایش دقت بازیابی می‌شود.

³¹ Google Translator

32 NIP

NLI

```
In [117]: M booksen.tags[0]
Out[117]: 'Classical literature. The boy's name was "Santiago". It was dusk when he arrived at an abandoned church with his sheep. The roof of the church collapsed a long time ago and a large plantain tree had grown in the place that was previously used for sacrificial vessels. He decided to spend the night there. He brought all the sheep into the sack through the broken door and block the door with some pole and board so that no sheep could leave during the night. There was no trace of a wolf in that area, but one night when one of his sheep was separated from the flock, he had to spend the whole next day looking for that animal. He swept the floor with his coat and lay down, placing the book he had finished a while ago on his pillow. "I must start reading thicker books," she said to herself, "because they will both take longer to read and be more comfortable pillows." سه یعنی پنجه کریمین Classical literature Classical literature'
```

```
In [118]: M #Stemming usually refers to a crude heuristic process that chops off the ends of words in the hope of achieving this goal correctly most of the time, and often includes
stemmer = SnowballStemmer("english")
def stem(text):
y = []
for i in text.split():
y.append(stemmer.stem(i))
return ' '.join(y)
booksen['tags'] = booksen['tags'].apply(stem)

```

```
In [119]: M booksen.tags[0]
Out[119]: 'classic literatur the boy name was "santiago". it was dusk when he arriv at an abandon church with his sheep. the roof of the church collaps a long time ago and a larg plantain tree had grown in the place that was previous use for sacr vessels. he decid to spend the night there. he brought all the sheep into the sack through the broken door and block the door with some pole and board so that no sheep could leav dure the night. there was no trace of a wolf in that area, but one night when one of his she ep was separ from the flock, he had to spend the whole next day look for that animal. he swept the floor with his coat and lay down, place the book he had finish a while ago on his pillow. "i must start read thicker books," she said to herself, "becaus they will both take longer to read and be more comfort pillows." اند گفه پنجه کریمین clas sic literatur classic literatur'
```

شکل ۱۶

۳- برای آماده کردن متن، عملیات مربوط به توکن بندی کاراکتر های اضافه را انجام می دهیم. یعنی فضاهای اضافی، کاما و ... متن را حذف کرده و آن را به رشته ای از کلمات تبدیل می کنیم. مثلا در متن زیر در شکل نشان داده شده نقطه و " " حذف شده است.

```
Out[119]: 'classic literatur the boy name was "santiago". it was dusk when he arriv at an abandon church with his sheep. the roof of the church collaps a long time ago and a larg plantain tree had grown in the place that was previous use for sacr vessels. he decid to spend the night there. he brought all the sheep into the sack through the broken door and block the door with some pole and board so that no sheep could leav dure the night. there was no trace of a wolf in that area, but one night when one of his she ep was separ from the flock, he had to spend the whole next day look for that animal. he swept the floor with his coat and lay down, place the book he had finish a while ago on hi s pillow i must start read thicker books she said to herself becaus they will both take longer to read and be more comfort pillows.' سه یعنی پنجه کریمین clas sic literatur classic literatur'
```

```
In [120]: M booksen['tags'] = booksen['tags'].str.replace('[\w\s]', '')
booksen['tags'][0]
```

```
Out[120]: 'classic literatur the boy name was "santiago". it was dusk when he arriv at an abandon church with his sheep the roof of the church collaps a long time ago and a larg plan tain tree had grown in the place that was previous use for sacr vessels he decid to spend the night there he brought all the sheep into the sack through the broken door and block the door with some pole and board so that no sheep could leav dure the night there was no trace of a wolf in that area but one night when one of his sheep was separ from the flock he had to spend the whole next day look for that animal he swept the floor with his coat and lay down place the book he had finish a while ago on hi s pillow i must start read thicker books she said to herself becaus they will both take longer to read and be more comfort pillows' سه یعنی پنجه کریمین classic literatur clas sic literatur'
```

شکل ۱۷

۴- از تکنیک تی اف آی دی اف برای وزن دار کردن کلمات و تبدیل آنها به برداری از اعداد استفاده می کنیم.

Term Frequency -Inverse Document Frequency [5] **TF-IDF** می باشد و به معنای فراوانی وزنی کلمه کلیدی است. **TF-IDF** صرفا میزان تکرار یک کلمه کلیدی یا عبارت را در صفحه نشان نمی دهد، بلکه هدف آن نشان دادن اهمیت کلمه کلیدی مورد نظر از طریق مقایسه تعداد تکرار کلمه در متن با تکرار آن کلمه در مجموعه ای بزرگ تراز مستندات (کل صفحات سایت) می باشد.

روش محاسبه **TF-IDF**

فرمول دو عبارت فوق به شرح زیر می باشد:

$TF = \text{عبارت از تقسیم تعداد تکرار کلمه بر تعداد کل کلمات محتوا}$

$IDF = \text{عبارت از لگاریتم تقسیم تعداد کل محتوا بر محتواهایی که شامل کلمه مورد نظر هستند.}$

(علت استفاده از لگاریتم در این فرمول این است که قصد داریم درون کل محتواهای سایت نرمال سازی انجام دهیم)

$$W_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

TF-IDF

Term x within document y

$tf_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents

به عنوان مثال اگر کلمه "سبد و سبد سهام" را با هم مقایسه نماییم در می‌یابیم که کلمه "سبد" بسیار بیشتر از "سبد سهام" مورد استفاده قرار می‌گیرید، بر مبنای تناوب تکرار این دو عبارت می‌توان نتیجه گرفت اگر عبارت "سبد سهام" در متن تکرار شود کلمه مهمی است در صورتی که آستانه مهم در نظر گرفتن کلمه "سبد" بسیار بالاتر است!

```
# TF-IDF is an abbreviation for Term Frequency Inverse Document Frequency.
tfidf = TfidfVectorizer(stop_words='english')
tfidf_matrix = tfidf.fit_transform(booksen['tags'])
tfidf_matrix.shape
```

شکل ۱۲

۵- تابع پیشنهاد دهنده کتاب را وارد می‌کنیم.

```
def get_recommendations(title):
    # Get the index of the book that matches the title
    idx = booksen.index[booksen['Name'] == title][0]
    # show given book image
    try:
        a=io.imread(f'https://img.ketabrah.ir/img/l/{booksen.loc[idx, "Images"]}')
        plt.imshow(a)
        plt.axis('off')
        plt.title(title)
        plt.show()
    except:pass
    print('Recommendations\n')

    # Get the pairwsie similarity scores of all books with that book
    sim_scores = list(enumerate(
        cosine_similarity(
            tfidf_matrix,
            tfidf_matrix[idx])))
    # Sort the books based on the similarity scores
    sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)

    # Get the scores of the 10 most similar books
    sim_scores = sim_scores[1:10]

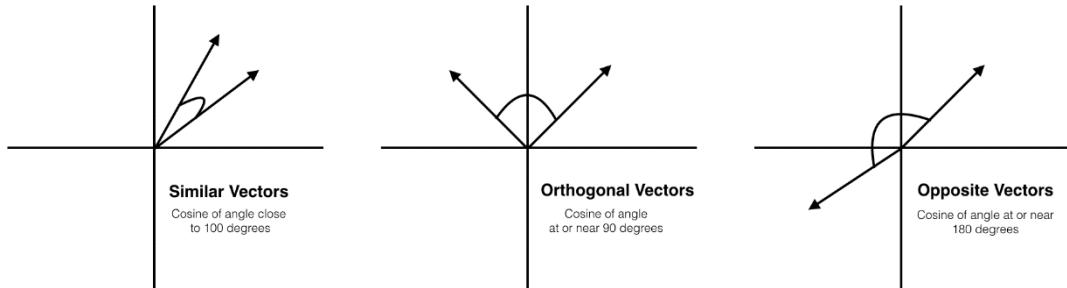
    # Get the book indices
    book_indices = [i[0] for i in sim_scores]

    # Return the top 10 most similar books
    result = booksen.iloc[book_indices]

    # show reco. book posters
    fig, ax = plt.subplots(2, 4, figsize=(15,15))
    ax=ax.flatten()
    for i, j in enumerate(result.Images):
        try:
            ax[i].axis('off')
            ax[i].set_title(result.iloc[i].Name)
            a = io.imread(f'https://img.ketabrah.ir/img/l/{j}')
            ax[i].imshow(a)
            ax[i].show()
        except: pass
    fig.tight_layout()
    fig.show()
```

شکل ۱۳

معیارهای شباخت، معیارهایی مانند معیارهای فاصله هستند که میزان دور و یا نزدیک بودن دو موجودیت را مشخص می‌کنند. بدیهی است که معیار شباخت با معیارهای فاصله رابطه عکس دارند و به عبارتی هر چه میزان شباخت بیشتر باشد می‌توان نتیجه گرفت فاصله‌ی دو شیء کمتر است. در این سیستم پیشنهاد دهنده از معیار شباخت کسینوسی^{۳۴} [6] استفاده شده است.



شکل ۲

در صورت انطباق دو بردار (در این معیار نشانه شباخت کامل است) که زاویه‌ی بین دو بردار صفر می‌باشد مقدار آن برابر ۱ خواهد شد و در کمترین میزان شباخت دو بردار یعنی اگر زاویه بین دو بردار $1/10^{\circ}$ درجه باشد نتیجه این معیار ۱ - خواهد شد.

$$\text{similarity}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

شکل ۳

ویژگی‌های شباخت کسینوسی

۱- شباخت کسینوسی به ضرایب مختلف ویژگی‌ها توجه نمی‌کند.

برای مثال فرض کنید فردی به فیلم‌های $m1$ و $m2$ حداقل امتیاز را با مقدار یک ثبت کرده و دیگری برای آن دو فیلم حداکثر امتیاز را با مقدار ۵ ثبت کرده است، در این شرایط شباخت کسینوسی این دو فرد برابر یک می‌شود که نشانه شباخت کامل در این معیار است.

۲- این معیار مقادیر صفر برای ویژگی‌ها در نظر نمی‌گیرد.

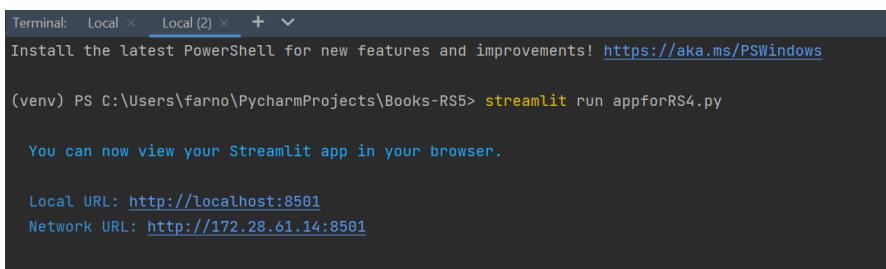
یعنی مناسب زمانیست که صفر به معنای خالی و نامشخص بودن حالت ویژگی در مسئله باشد. مثلاً اگر دو مقدار صفر نشان دهنده‌ی ثبت نشدن امتیاز دو فرد برای یک محصول باشد می‌توان از این معیار استفاده نمود، ولی اگر نشان دهنده‌ی امتیاز پایین آن دو شخص باشد که در این حالت می‌توان نتیجه گرفت که نظر این دو فرد مشابه است این معیار به خوبی عمل نمی‌کند زیرا این حالت را در نظر نمی‌گیرد. در چنین شرایطی جایجایی بازه‌ی امتیازات برای مثال بجای 0 تا 1 به یک تا 6 می‌تواند یکی از راه کارهای مفید باشد تا بتوان از این معیار شباخت کسینوسی استفاده نمود. سپس در محیط پایی چارم^{۳۵} و با استفاده از کتابخانه استریم لیت^{۳۶} سیستم پیشنهاد دهنده به یک محیط

³⁴ Cosine Similarity

³⁵ pycharm

³⁶ Streamlit

وب ملی^{۳۷} برای سادگی کار کاربر انتقال داده شد. استریم لیت، فریمورکی متن باز است که به شما اجازه می‌دهد اپلیکیشن‌ها و مدل‌های علم‌داده‌تان را به راحتی توسعه دهید و پروژه‌هایتان را با دیگران به اشتراک بگذارید. به کمک این فریمورک جذاب قادر خواهید بود داشبوردهای زیبا و تعاملی بسازید و در عین حال مدل‌های خود را توسعه دهید.



```
Terminal: Local × Local (2) × + 
Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

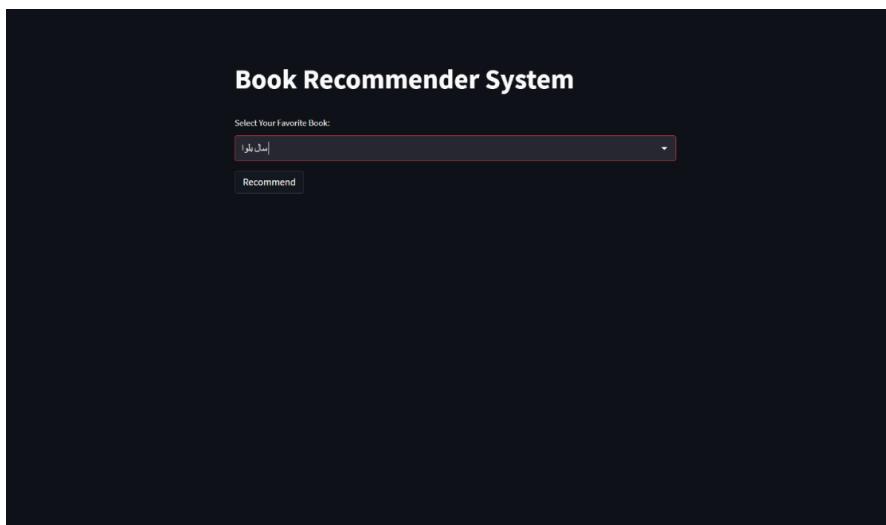
(venv) PS C:\Users\farno\PycharmProjects\Books-RS5> streamlit run appforRS4.py

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://172.28.61.14:8501
```

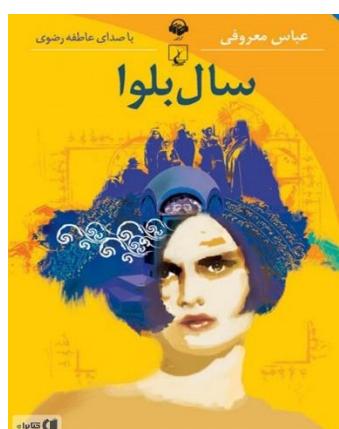
شکل ۳۲

صفحه نمایش به شکل زیر می‌باشد:



شکل ۳۳

سپس با انتخاب کتاب مورد علاقه از لیست و زدن گزینه *Recommend*^{۳۸}، کتاب مشابه به شما پیشنهاد می‌شود. که با سایت کتابراه و توضیحات این سایت تطابق داده شده است.

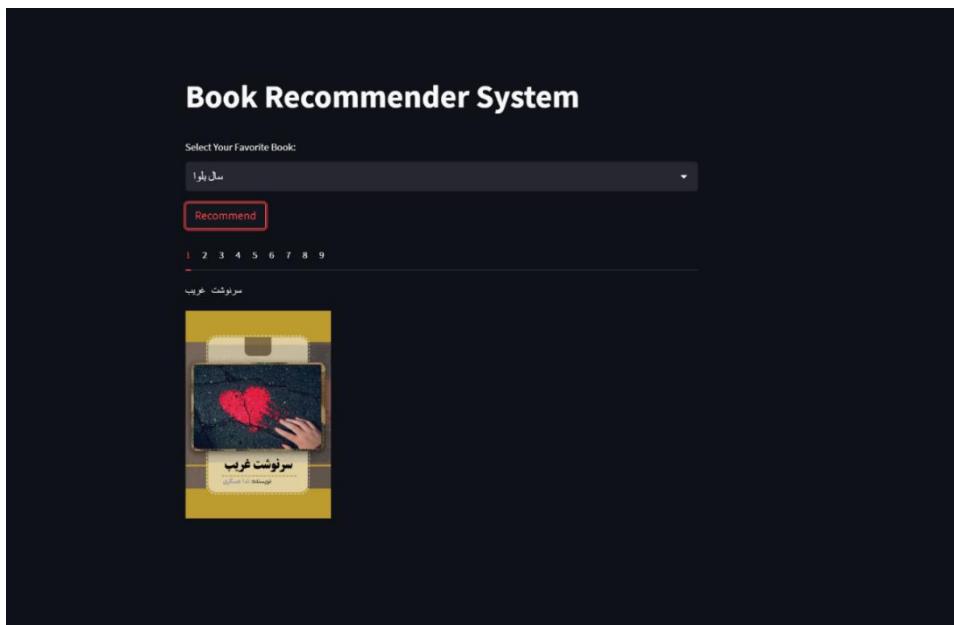


³⁷ Local Host

معرفی کتاب صوتی سال بلوا

کتاب صوتی سال بلوا اثر عباس معروفی روایتی از هفت شب است که شما را سال‌ها عقب برده و تداعی‌گر کل تاریخ ایران و زندگی غمانگیز نوشافرین است. نوشافرینی که خود نمادی از مظلومیت تمامی زنان و دختران رنج کشیده است.

کتاب سال بلوا (Sal-e- Balva) به سبک سیال ذهن نگاشته شده است. خانواده‌ای که با امید و آرزو نقل مکان می‌کنند اما بلوایی از جانب مردم، نقشه‌هایشان را نقش برآب می‌کنند. داستانی که به صورت اول شخص اتفاق می‌افتد و نویسنده از پرداخت به حاشیه پرهیز می‌کند و مخاطب را مستقیم وارد داستان می‌کند. ما با زنی رویه‌رو هستیم که در زمان حال به سر می‌برد و بعد از رویه‌رو شدن با واقعیت، دوباره به خاطرات خود باز می‌گردد. اوست که بخش‌هایی از کتاب را روایت می‌کند و بخش‌هایی دیگر کتاب توسط دانای کل روایت می‌شود.



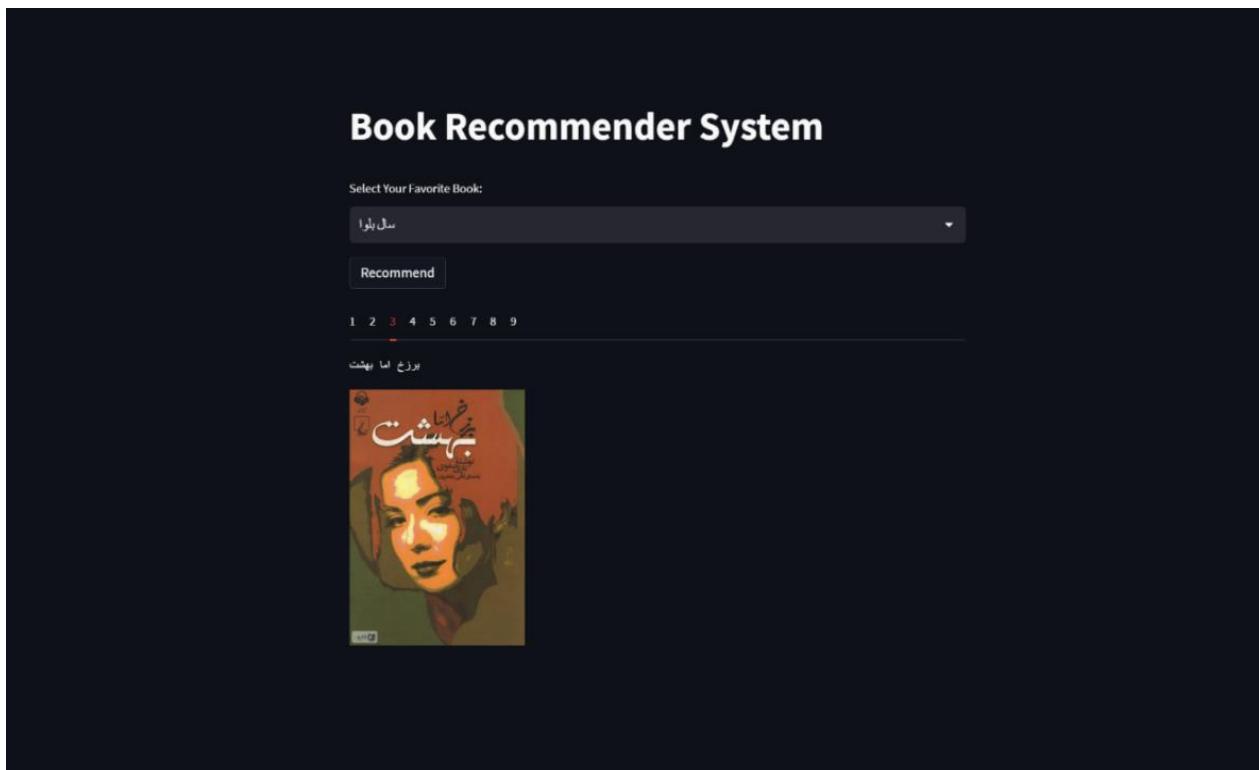
شکل ۴۳

معرفی کتاب سرنوشت غریب

کتاب سرنوشت غریب به قلم ندا عسگری، داستان زندگی دختری به نام منا را به تصویر می‌کشد که در پی داستان عاشقانه‌اش با پسری به نام امیر، روزگار دوران سختی را پیش رویش می‌گذارد.

در بخشی از کتاب سرنوشت غریب می‌خوانیم:

با قدم‌هایی لزان در کوچه به راه افتادم. گویی اینجا برایم ته دنیا بود. مثل دختر بچه‌ی بی‌پناهی بودم که احساس غربت و ترس تمام وجودش را گرفته. بلافصله خودم را به ترمینال رساندم. دلم نمی‌خواست بیشتر از این، در این شهر غریب بمانم. بليط خريدم و روی يكى از صندلی‌های سالن ترمینال منتظر ماندم.



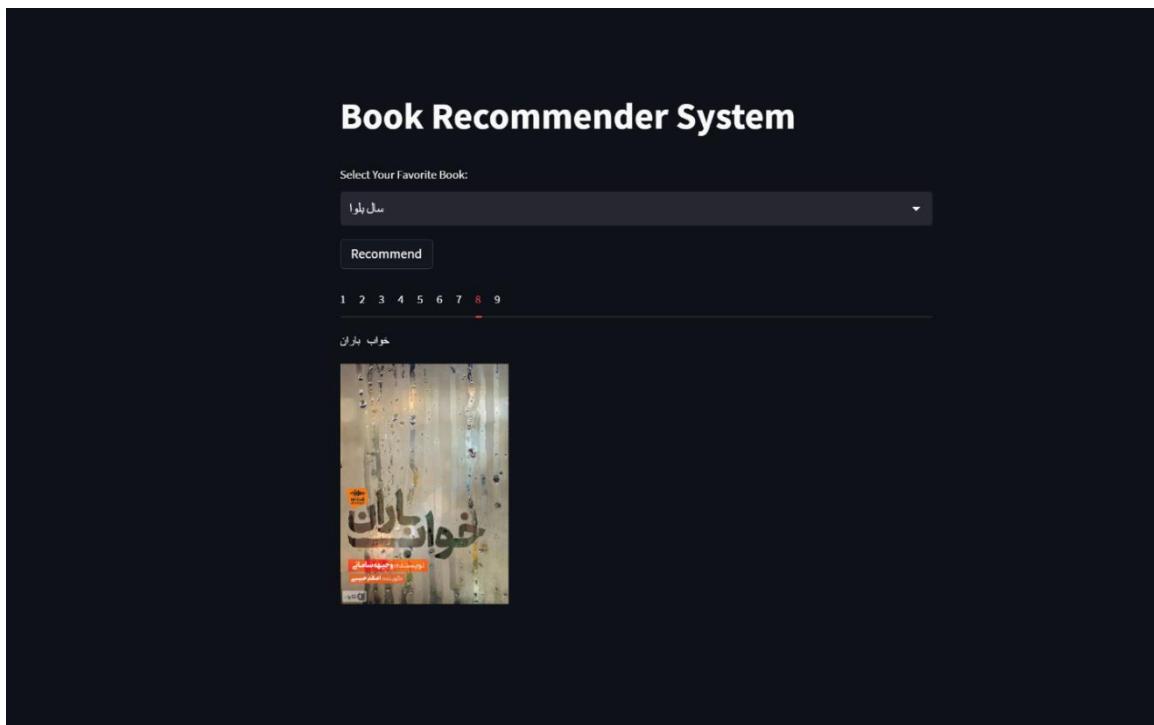
شکل ۱۵

معرفی کتاب صوتی برزخ اما بهشت

نازی صفوی در کتاب صوتی برزخ اما بهشت، از عشق با محوریت زن و رنجها و سختیهای آن سخن می‌گوید.

برزخ اما بهشت تصویری از زندگی زنی به نام ماهنوش است با روحی زخم خورده، که پس از پایان دادن به زندگی مشترک سراسر رنجبارش، به نزد خانواده‌اش باز می‌گردد تا در آن مامن مالوف، آلام گذشته‌اش را به فراموشی بسپارد. غافل از این که تنها‌یی زن، رنج مضاعفی است که بر شمار زخم‌های روحش می‌افزاید. با این همه، این بار، روزگار جفاکار گویی بر آن است تا آرامش از دست رفته را با حضور عزیزانش به او بازگرداند.

در این هنگام دخترخاله‌ی ماهنوش، رعنا که حتی نسبت به خواهرهایش بیشتر احساس صمیمیت و راحتی می‌کند همراه با دخترش کیمیا به ایران می‌آید و وضعیت ماهنوش رو به بهبودی می‌رود اما مدتی بعد مشخص می‌شود که رعنای سرطان دارد و بعد از مرگ رعنای، ماهنوش سرپرستی کیمیا، دختر رعنای را به عهده می‌گیرد، در این راه حسام، برادر رعنای و پسرخاله‌ی ماهنوش به او کمک می‌کند و....

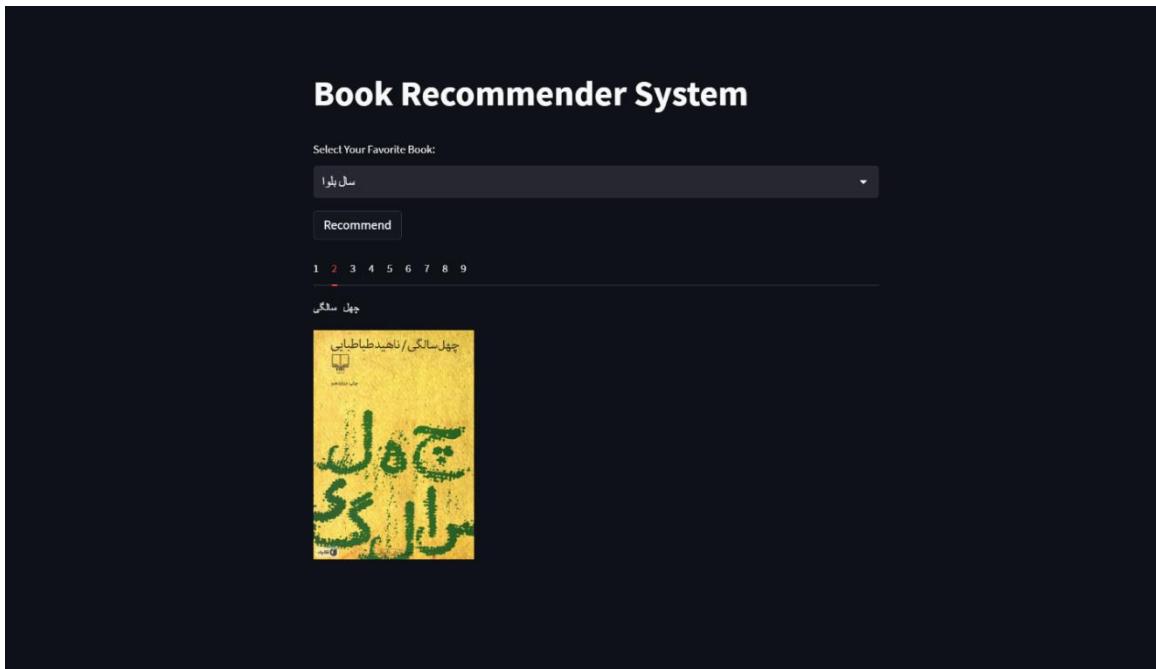


شکل ۱۵

معرفی کتاب صوتی خواب باران

سرنوشت بازی عجیبی دارد! **وجیهه سامانی** در کتاب صوتی خواب باران در پس زمینه داستانی عاشقانه نگاهی واقع‌بینانه به موضوعات اخلاقی و اجتماعی می‌اندازد؛ به معضلات روز جوانان مثل فقر فرهنگی، فقر اقتصادی، اعتیاد، خودکشی و ازدواج.

قرار گرفتن در تنگناها و مشکلات زندگی، دختری را به سمت خودکشی می‌کشاند؛ داستان از جایی شروع می‌شود که خانواده‌ای مهربان این دختر را نجات می‌دهند و او دوباره به زندگی برمی‌گردد. این کتاب تعریفی صحیح و واقعی از سرنوشت و قضا و قدر ارائه می‌کند و مرز باریک میان تردید و ایمان را به درستی نمایش می‌دهد.



شکل ۱۷

معرفی کتاب صوتی چهل سالگی

کتاب صوتی چهل سالگی نوشته ناهید طباطبایی، درباره‌ی عشق، جوانی، خانواده و موسیقیست و روایت‌گر حسرت‌های زندگی یک زن چهل ساله است.

داستان کتاب چهل سالگی درباره‌ی زنی به نام آلاله است که در مرز چهل سالگی قرار گرفته و دچار دگرگونی شده است. او همراه با همسری بسیار منطقی و دختری شاد زندگی می‌کند. او زندگی آرام و شغل مناسبی دارد. آلاله مسئول کنسرت‌گذاری در تالاری بزرگ است و رهبر یکی از کنسرت‌ها قرار است که از اروپا به ایران بیاید و او متوجه می‌شود که آن فرد کسی نیست جز عشق دوران جوانی‌اش. او شکه می‌شود، یاد گذشته می‌افتد، یاد عاشقی‌های گذشته و نمی‌داند که با او چگونه برخورد کند. آلاله دیگر آن دختر جوان نیست و نزدیک شدن دوران پیری دلهزه به جان او اندادخته است.

پیوست

کدهای ساخت نرم افزار در وبسایت استریمیت

داشبورد مصورسازی داده در پاوربی آی

دیتابست انگلیسی برای سیستم پیشنهاد دهنده

دیتابست سایت کتابراه برای پیش پردازش و مصورسازی داده

کدهای پایتون پیش پردازش و مصورسازی داده

کدهای پایتون سیستم پیشنهاد دهنده

راهنمای استفاده از سیستم پیشنهاد دهنده

منابع

- [1] "سایت کتابراه،" [متصل].
- [2] "کتوپارس،" Available: https://www.octoparse.com/?gclid=Cj0KCQiAi8KfBhCuARIsADp-A556HQoh2IH70Spmc-JkLPVmsy1205kNHQ5aRj8O6XEdWzzs74_roWgaArhVEALw_wcB. [متصل].
- [3] J. Brownlee ,'"Why One-Hot Encode Data in Machine Learning?". Machinelearningmastery." 2017 .Available:
<https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/>
- [4] F. Ricci ،L. Rokach ،B. Shapira ,'"Recommender Systems Handbook (3 ed.).," New York ،Springer ،2022 ،pp. 1-35.
- [5] A. Rajaraman و J. Ullman ,'"Mining of Massive Datasets.." 2011 ،pp. 1-17.
- [6] A. Singhal ,'"Modern Information Retrieval: A Brief Overview ،" تأليف "Bulletin of the IEEE Computer Society Technical Committee on Data Engineering ،" 2001 ،pp. 35-43.