

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



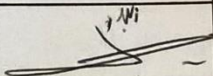
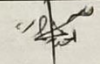

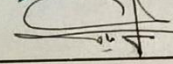
دانشکده

مهندسی صنایع و سیستم ها

بسمه تعالی  
تاییدیه اعضای هیات داوران

آقای رسول نوروزی نیکجه پایان نامه خود را با عنوان: طراحی یک سامانه پیشنهادگر مکان های گردشگری (مورد مطالعاتی: استان تهران) در تاریخ ۱۳۹۷/۰۶/۲۵ ارائه کردند.

اعضای هیات داوران نسخه نهایی پایان نامه را از نظر فرم و محتوا بررسی نموده و پذیرش آنرا برای تکمیل درجه کارشناسی ارشد تایید می نمایند.

اعضای هیات داوران	نام و نام خانوادگی	رتبه علمی	امضاء
استاد راهنما	دکتر امیر البدوی	استاد	
استاد مشاور	دکتر الهام آخوندزاده نوقایی	استادیار	
استاد ناظر داخلی	دکتر الهام یآوری	استادیار	
استاد ناظر خارجی	دکتر سمیه علیزاده	استادیار	
نماینده شورای تحصیلات تکمیلی	دکتر الهام یآوری	استادیار	

## دستورالعمل حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی دانشگاه تربیت مدرس

**مقدمه:** با عنایت به سیاستهای پژوهشی دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیات علمی، دانشجویان، دانش آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان نامه، رساله و طرحهای تحقیقاتی که با هماهنگی دانشگاه انجام شده است، موارد ذیل را رعایت نمایند:

**ماده ۱- حقوق مادی و معنوی پایان نامه ها / رساله های مصوب دانشگاه متعلق به دانشگاه است و هرگونه بهره برداری از آن باید با ذکر نام دانشگاه و رعایت آیین نامه ها و دستورالعمل های مصوب دانشگاه باشد.**

**ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان نامه / رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و استاد راهنما مسئول مکاتبات مقاله باشد.**

**تبصره:** در مقالاتی که پس از دانش آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان نامه / رساله نیز منتشر می شود نیز باید نام دانشگاه درج شود.

**ماده ۳- انتشار کتاب حاصل از نتایج پایان نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با مجوز کتبی صادره از طریق حوزه پژوهشی دانشگاه و بر اساس آئین نامه های مصوب انجام می شود.**

**ماده ۴- ثبت اختراع و تدوین دانش فنی و یا ارائه در جشنواره های ملی، منطقه ای و بین المللی که حاصل نتایج مستخرج از پایان نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق حوزه پژوهشی دانشگاه انجام گیرد.**

**ماده ۵- این دستورالعمل در ۵ ماده و یک تبصره در تاریخ ۱۳۸۴/۴/۲۵ در شورای پژوهشی دانشگاه به تصویب رسیده و از تاریخ تصویب لازم الاجرا است و هرگونه تخلف از مفاد این دستورالعمل، از طریق مراجع قانونی قابل پیگیری می شود.**

نام و نام خانوادگی

امضاء

رسول نوروزی



### آیین نامه چاپ پایان نامه (رساله) های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله) های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله) ی خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثار علمی» دانشگاه اطلاع دهد.

ماده ۲: در صفحه سوم کتاب (پس از برگ شناسنامه) عبارت ذیل را چاپ کند:

«کتاب حاضر، حاصل پایان نامه کارشناسی ارشد/ رساله دکتری نگارنده در رشته

در دانشکده

، مشاوره سرکار خانم/جناب آقای دکتر

از آن دفاع شده است.»

خانم/جناب آقای دکتر

و مشاوره سرکار خانم/جناب آقای دکتر

ماده ۳: به منظور جبران بخشی از هزینه های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به «دفتر نشر آثار علمی» دانشگاه اهدا کند. دانشگاه می تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.

ماده ۴: در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده را به عنوان خسارت به دانشگاه تربیت مدرس، تأدیه کند.

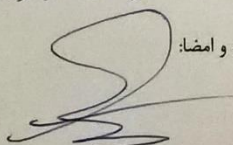
ماده ۵: دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت بهای خسارت، دانشگاه می تواند خسارت مذکور را از طریق مراجع قضایی مطالبه و وصول کند؛ به علاوه به دانشگاه حق می دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶: اینجانب رسول نوروزی دانشجوی رشته مهندسی فناوری اطلاعات مقطع کارشناسی ارشد

تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی: رسول نوروزی

تاریخ و امضا:





## گزارش پایان نامه

طراحی سامانه پیشنهادگر مکان‌های گردشگری (مورد مطالعاتی: استان

تهران)

دانشجو

رسول نوروزی

استاد راهنما

دکتر امیر البدوی

استاد مشاور

دکتر الهام آخوندزاده

تابستان ۱۳۹۷

این پایان نامه را ضمن تشکر و سپاس بیکران و در کمال افتخار و امتنان تقدیم می نمایم به:

محضر ارزشمند پدر و مادر عزیزم به خاطر همه ی تلاشهای محبت آمیز ی که در دوران مختلف زندگی ام انجام داده اند و بامهربانی چگونه زیستن را به من آموخته اند.

از استاد گرامیم جناب آقای دکتر امیر البدوی بسیار سپاسگذارم چرا که بدون راهنمایی‌های ایشان تامین این پایان نامه بسیار مشکل می‌نمود.

از مشاور گرامی سرکار خانم دکتر الهام آخوندزاده به دلیل یاری‌ها و راهنمایی‌های بی‌چشمداشت ایشان که بسیاری از سختی‌ها را برایم آسان‌تر نمودند، سپاسگزارم.





## چکیده

گردشگری و تفریح نقش بسیار پر رنگی در زندگی انسان دارد و در این میان فناوری اطلاعات در آمیختگی با صنعت گردشگری موجب مزیت دو سویه هم برای کاربران و هم ارائه دهندگان خدمات تفریحی و گردشگری شده است. از سوی دیگر فناوری موجب افزایش بی رویه اطلاعات در وب شده که این امر منجر به افزونگی اطلاعات شده است و در نتیجه کاربر جهت دریافت اطلاعات مناسب برای تصمیم گیری دچار مشکل می شود و فقدان دانش دامنه موجب شده کاربر نتواند خواسته های خود را به روشنی در موتورهای جستجو بیان کند. برای غلبه بر این مشکل از سامانه های پیشنهادگر استفاده می شود که از فن ها و رویکردهای متنوعی استفاده می کنند. سامانه های پیشنهادگر مبتنی بر آنتولوژی نیاز به قضاوت خبرگان برای ارتباط سطوح آنتولوژی دارند. از طرفی روش های رایج پیشنهادگرها مانند خوشه بندی ممکن است برای هر داده ای متناسب نباشد و خوشه های مناسبی ارائه ندهند و همچنین پیشنهادات ارائه شده از تنوع مناسبی برخوردار نباشد، از سوی دیگر جهت رفع مشکل شروع سرد دریافت اطلاعات صریح دموگرافیک از کاربر می تواند برای کاربران ناخوشایند باشد. در این پژوهش که استان تهران به عنوان مطالعه موردی انتخاب شده است، گروه های سطح دوم آنتولوژی به کمک تحلیل عاملی بدست آمده و روشی ارائه شده که با استفاده از رویکرد دسته بندی های چندبرچسب ارتباط بین سطح گروه ها و سطح مفهوم آنتولوژی را بدون قضاوت انسانی برقرار می کند. در روش ارائه شده در این پژوهش دریافت اطلاعات از کاربر شامل اطلاعات دموگرافیک نمی باشد و صرفاً علاقه مندی کاربر به گروه های گردشگری سنجیده می شود. نتایج به صورت برون خطی، مبتنی بر سناریو و مطالعه کاربری ارزیابی شد و در همه موارد پاسخگوی اهداف پژوهش بود.

**کلیدواژه ها:** سامانه پیشنهادگر، گردشگری، دسته بندی های چندبرچسب، آنتولوژی، تحلیل عاملی، کلیشه ها

## فهرست مطالب

عنوان	شماره صفحه
<b>فصل اول</b>	<b>۱</b>
۱,۱ مقدمه	۲
۱,۲ تعریف مسئله و ضرورت پرداخت به آن	۳
۱,۳ اهداف پژوهش	۶
۱,۴ سوالات پژوهش	۶
۱,۵ جنبه‌های نوآوری پژوهش	۷
۱,۶ مواد و روش انجام تحقیق	۷
۱,۷ ساختار پژوهش	۸
<b>فصل دوم</b>	<b>۹</b>
۲,۱ مقدمه	۱۰
۲,۲ سامانه‌های اولیه جستجوی اطلاعات	۱۱
۲,۲,۱ موتورهای جستجو	۱۱
۲,۲,۲ سامانه‌های یاریگر	۱۱
۲,۲,۳ سامانه‌های پالایش اطلاعات و سامانه‌های بازیابی	۱۱
۲,۳ سامانه پیشنهادگر	۱۲
۲,۳,۱ تعاریف اولیه در سامانه‌های پیشنهادگر	۱۳
۲,۳,۲ فرآیند شخصی‌سازی اطلاعات	۱۵
۲,۳,۳ مزایا و معایب پیشنهادگرها	۱۶
۲,۳,۴ رویکردها و فن‌های مورد استفاده در پیشنهادگرها	۱۷
۲,۳,۵ روش‌های ارزیابی سامانه‌های پیشنهادگر و بسندگی آن‌ها	۲۵

۲۸.....	۲,۳,۶ مشکلات و چالش‌های سامانه‌های پیشنهادگر
۲۹.....	۲,۴ پیشنهادگرها و گردشگری
۳۱.....	۲,۴,۱ نقش‌ها و شخصیت‌ها در پیشنهادگرهای گردشگری
۳۳.....	۲,۴,۲ مروری بر پژوهش‌های پیشین پیشنهادگرهای گردشگری
۳۷.....	۲,۵ فن‌ها و روش‌ها
۳۸.....	۲,۵,۱ یادگیری ماشین
۳۹.....	تحلیل عاملی
۴۲.....	۲,۵,۲ مدل‌های دسته‌بند
۴۷.....	۲,۵,۳ بررسی عملکرد مدل
۴۸.....	۲,۵,۴ دسته‌بندی چند برجسب
۵۱.....	۲,۶ اختصارات
۵۲.....	۲,۷ خلاصه فصل و نتیجه‌گیری
۵۳.....	<b>فصل سوم</b>
۵۴.....	۳,۱ مقدمه
۵۶.....	۳,۲ قدمها پژوهش
۵۶.....	۳,۲,۱ شناسایی فعالیت‌های گردشگری
۵۶.....	۳,۲,۲ جمع‌آوری اطلاعات پرسشنامه ۱
۵۶.....	۳,۲,۳ بررسی پایایی پرسشنامه ۱
۵۷.....	۳,۲,۴ تحلیل عاملی
۵۷.....	۳,۲,۵ ساخت آنتولوژی
۵۷.....	۳,۲,۶ جمع‌آوری اطلاعات پرسشنامه ۲
۵۸.....	۳,۲,۷ بررسی پایایی پرسشنامه ۲
۵۸.....	۳,۲,۸ پیاده‌سازی الگوریتم‌های دسته‌بند چند برجسب
۵۹.....	۳,۲,۹ ارزیابی برون خطی
۶۱.....	۳,۲,۱۰ تشریح عملکرد دسته‌بند
۶۱.....	۳,۲,۱۱ ارزیابی مبتنی بر سناریو
۶۱.....	۳,۲,۱۲ ارزیابی مطالعه کاربر
۶۱.....	۳,۲,۱۳ نتیجه‌گیری
۶۲.....	۳,۳ خلاصه و جمع‌بندی فصل: تشریح کلی عملکرد سامانه
۶۳.....	<b>فصل چهارم تجزیه و تحلیل یافته‌های پژوهش</b>

۶۴	۴,۱ مقدمه .....
۶۴	۴,۲ شناسایی فعالیت‌های گردشگری .....
۶۵	۴,۳ جمع‌آوری اطلاعات (پرسشنامه ۱) .....
۶۷	۴,۴ تحلیل عاملی پرسشنامه .....
۷۲	۴,۵ ساخت آنتولوژی .....
۷۳	۴,۶ جمع‌آوری اطلاعات (پرسشنامه ۲) .....
۷۵	۴,۷ ارزیابی برونخطی .....
۷۸	۴,۸ تشریح عملکرد دسته‌بند .....
۷۹	۴,۹ ارزیابی به صورت مبتنی بر سناریو .....
۸۰	۴,۱۰ ارزیابی مطالعه کاربر .....
۸۱	۴,۱۱ خلاصه فصل .....
۸۳	<b>فصل پنجم .....</b>
۸۴	۵,۱ مقدمه .....
۸۴	۵,۲ مرور و تحلیلی بر فصل‌های گذشته .....
۸۶	۵,۳ نتیجه‌گیری پژوهش .....
۸۸	۵,۴ محدودیتها و پیشنهادات آتی .....
۸۹	<b>مراجع .....</b>
۹۰	مراجع فارسی .....
۹۰	مراجع انگلیسی .....
۹۵	<b>پیوست‌ها .....</b>
۹۶	پیوست ((الف)) پرسشنامه اول .....
۹۸	پیوست ((ب)) پرسشنامه دوم .....

## فهرست جدول‌ها

عنوان	شماره صفحه
جدول ۱-۲ ماتریس نرخ دهی کاربران برای پیشنهاد اقلام .....	۱۳
جدول ۲-۲ نمونه‌هایی از روش‌های جمع‌آوری اطلاعات و شناخت کاربران .....	۱۵
جدول ۲-۳ نقش‌های گردشگری هفده گانه .....	۳۳
جدول ۲-۴ مروری بر پژوهش‌های گذشته در پیشنهادگرهای گردشگری .....	۳۶
جدول ۲-۵ اختصارات پرکاربرد در پیشنهادگرهای گردشگری و اختصارات استفاده شده در این پژوهش .....	۵۱
جدول ۴-۱ فهرستی از فعالیت‌های شناسایی شده و معادل/اختصار آن‌ها به زبان انگلیسی .....	۶۵
جدول ۲-۴ نتیجه آزمون بارتلت و کیسر .....	۶۸
جدول ۳-۴ واریانس‌های تشریح شده جهت انتخاب تعداد عامل مناسب .....	۶۸
جدول ۴-۴ جدول عامل‌ها به همراه بار عاملی برای هر متغیر قبل از چرخش واریمکس .....	۷۰
جدول ۵-۴ جدول عامل‌ها به همراه بار عاملی برای هر متغیر بعد از چرخش واریمکس .....	۷۱
جدول ۶-۴ جدول نام‌گذاری بر روی عامل‌های حاصل از تحلیل عاملی .....	۷۲
جدول ۷-۰ مقایسه عددی مدل‌های دسته‌بندی جهت پیش‌بینی فعالیت مورد نظر افراد .....	۷۶
جدول ۴-۸ نمایش عددی گروه‌های تاثیر گزار در پیش‌بینی فعالیت‌ها توسط مدل‌های دسته‌بندی کت بوست .....	۷۸

## فهرست شکل ها و نمودارها

عنوان	شماره صفحه
شکل ۱-۲ دسته بندی رویکردی پیشنهادگرها	۱۸
شکل ۲-۲ مقایسه فراوانی مقالات نوع پیشنهادگرهای گردشگری بر اساس عملکرد	۳۱
شکل ۲-۳ یک درخت ساده تصمیم	۴۴
شکل ۲-۴ نحوه عملکرد الگوریتم SVM	۴۵
شکل ۲-۵ نمایشی از عملکرد الگوریتم کا-نزدیکترین همسایه	۴۶
شکل ۲-۶ مثالی از داده های چند برچسب	۴۹
شکل ۲-۷ نمایش عملکرد روش باینری رلونس	۴۹
شکل ۲-۸ نمایشی از عملکرد روش دسته بندی زنجیره ای	۵۰
شکل ۲-۹ نمایشی از عملکرد روش لیبل پاورست	۵۰
شکل ۳-۱ نمای کلی از مراحل پژوهش	۵۵
شکل ۳-۲ نمایی از آنتولوژی مورد نظر پژوهش	۵۸
شکل ۳-۳ تصویر کلی از عملکرد سامانه پیشنهادگر	۶۲
شکل ۱-۴ نمودار نسبت جنسیتی پاسخ دهندگان به پرسشنامه	۶۶
شکل ۲-۴ نمودار نمایش درصدی بازه سنی افراد به پرسشنامه اول	۶۶
شکل ۳-۳ نمودار گرمایی همبستگی فعالیت های گردشگری شناسایی شده	۶۷
شکل ۴-۴ نمودار سنگ ریزه برای انتخاب تعداد عامل مناسب	۶۹
شکل ۴-۵ بخشی از آنتولوژی رسم شده برای پژوهش	۷۳
شکل ۴-۶ نمودار گرمایی ضریب همبستگی پیرسون بین گروه ها و فعالیت ها گردشگری	۷۴
شکل ۴-۷ نمودار شکل توازن دسته ها برای هر کدام از فعالیت ها	۷۵
شکل ۴-۸ نمودار مقایسه عملکرد دسته بندها جهت پیش بینی فعالیت مورد نظر کاربران	۷۷
شکل ۴-۹ نتایج پاسخ به پرسش ها در ارزیابی مطالعات کاربر	۸۱





# **فصل اول**

## **کلیات پژوهش**

## ۱،۱ مقدمه

انسان امروزی به مراتب بیشتر از اسلاف خود تغییر و جابه‌جایی را در زندگی خود تجربه می‌کند. به صورت کلی، تغییر و جابه‌جایی به دو شکل، عمودی و افقی صورت می‌گیرد. تحرک عمودی از تغییر پایگاه اجتماعی - اقتصادی انسان حکایت دارد؛ تحرک افقی خود به سه قسمت، تحرک فیزیکی ساده، مهاجرت و گردشگری تقسیم می‌شود. گردشگری به اشکال مختلفی صورت می‌گیرد، که مهمترین آنها سیاحت داخلی یا درون‌مرزی، و سیاحت خارجی یا برون‌مرزی و بین‌المللی است. گردشگری دیرباز در ایران رونق داشته و شواهد تاریخی فراوانی دال بر این مدعا وجود دارد، اما، صنعت گردشگری در ایران، آنچنان که باید، رشد و توسعه نیافته است و در نتیجه موفقیت‌چندانی برای جذب گردشگر و همچنین دستیابی به مزایای اقتصادی، فرهنگی، اجتماعی، و سیاسی صورت نگرفته است (میثم ۱۳۹۱، et al.).

در حال حاضر گردشگری به عنوان یکی از بزرگترین و متنوع‌ترین صنایع دنیا مطرح می‌باشد و رشد سریع آن تغییرات اجتماعی، اقتصادی و محیطی فراوانی را به دنبال داشته است و به همین دلیل نیز به یک حوزه مهم مطالعاتی بین پژوهشگران مبدل شده است. از نظر اقتصادی، گردشگری بین‌المللی بیشترین عایدی را ایجاد می‌کند، دریافته‌های ارزی حاصل از گردشگری بین‌المللی بیشتر از محصولات نفتی، خودرو و تجهیزات ارتباطی تأمین می‌شود. گردشگری نقش مهمی در ترغیب سرمایه‌گذاری در زیرساخت‌ها، ایجاد درآمد برای دولت و اشتغالزایی مستقیم و غیر مستقیم در سراسر دنیا داشته است (دکتر مهدی and دکتر علی قلی پور، ۱۳۸۸).

امروزه نقش و تأثیر فناوری اطلاعات و ارتباطات و صنعت گردشگری بر رشد اقتصادی کشورها بر کسی پوشیده نیست. با بهره‌گیری از فضای اینترنت و به‌وجود آمدن گردشگری الکترونیکی، فناوری اطلاعات و ارتباطات، یکی از عناصر اساسی گردشگری به‌شمار رفته و کارایی این صنعت را افزایش داده است. چهار ویژگی مهم را می‌توان به عنوان دلایل تجویز گسترش استفاده از فناوری اطلاعات، برای رونق صنعت

گردشگری مطرح نمود: نخستین ویژگی، پایین بودن متوسط سرمایه لازم برای ایجاد هر شغل در حوزه‌های مرتبط با فناوری اطلاعات در این صنعت است. به‌ویژه با توجه به کمبود شدید منابع سرمایه‌گذاری در کشور از یکسو و زیاد بودن تعداد متقاضیان اشتغال از سوی دیگر، این ویژگی بسیار اهمیت میابد. در واقع، گسترش شغل‌های مرتبط با اینترنت، باعث می‌شوند تا بدون نیاز به سرمایه‌گذاری‌های هنگفت، بتوان تعداد قابل توجهی شغل جدید در این صنعت ایجاد نمود و چنین مسئله‌ای به‌خصوص برای کشورهایی که با کمبود منابع مالی برای سرمایه‌گذاری در این بخش مواجه هستند، می‌تواند بسیار با اهمیت تلقی گردد. ویژگی دوم این است که با رشد سریع تقاضا در سطح دنیا برای خدمات مرتبط با فناوری اطلاعات از قبیل تبلیغات، بازاریابی، برنامه‌نویسی، خدمات امنیت شبکه و تولید محتوا برای سایت‌های اینترنتی در گردشگری، گسترش فناوری اطلاعات و ارتباطات در این بخش ضروری به نظر می‌رسد. ویژگی مهم دیگری که باعث تأثیر مثبت توسعه فناوری اطلاعات روی صنعت گردشگری می‌شود، عبارت است از: نقش مهم فناوری اطلاعات و ارتباطات در کاهش هزینه توسط تعداد زیادی از بنگاه‌های اقتصادی و ادارات دولتی در ارائه خدمات به گردشگران؛ مثلاً از طریق خدمات الکترونیکی، متوسط زمان لازم برای حضور آنها در محل کار کاهش می‌یابد. چنین روندی منجر به صرفه‌جویی در هزینه‌های جانبی ناشی از حضور آنها در محل کار می‌شود. آخرین کانال تأثیرگذاری توسعه فناوری اطلاعات و ارتباطات بر صنعت گردشگری، به تأثیر فناوری اطلاعات و ارتباطات در جذب گردشگر است. در دنیای کنونی، بیشتر گردشگران به‌دلیل اینکه سامانه سنتی زمان و هزینه زیادی صرف می‌کرد، از طریق جستجوی اینترنتی به تعیین محل گردش خود می‌پردازند. در تحقیقی ارائه شده از جمی در سال ۲۰۰۵، عنوان شده است که ۶۹ درصد از کره‌ای‌ها، ۶۵ درصد از انگلیسی‌ها و چینی‌ها، ۶۴ درصد از فرانسوی‌ها و ۶۲ درصد از ایتالیایی‌ها از اینترنت مکان گردشگری خود را انتخاب می‌کنند. پور فرج و همکاران از قول ورنر بیان می‌کنند که ۸۰ درصد از هتل‌های اروپا و بیش از ۹۰ درصد از هتل‌ها در استرالیا دارای وبسایت هستند و ۶۳ درصد از آنها خرید الکترونیکی و ۷۳ درصد از آنها به‌صورت برخط نیازمندی‌های مشتریان خود را برآورده می‌کنند (علیرضا et al., ۱۳۸۷).

## ۱.۲ تعریف مسئله و ضرورت پرداخت به آن

صنعت گردشگری از اینترنت به‌عنوان ابزاری برای ارائه خدمات به گردشگران استفاده می‌کند و از سویی دیگر گردشگران اطلاعات موردنیاز سفر و رزرو مقاصد گردشگری خود را از طریق فناوری وب انجام می‌دهند که این امر موجب مزیت دوسویه برای هر دو طرف در این صنعت شده است. بنا به (Neidhardt et al., ۲۰۱۵) وب اکنون بخشی جدایی‌ناپذیر سفر و صنعت گردشگری است و تأثیرگذاری آن بسیار

بیشتر از پیش‌بینی که در سال‌های ۱۹۹۰ زمانی که e-tourism شروع به پیدایش کرد، است. بنا به گزارش کمیسیون گردشگری اروپا بیش از دوسوم مردم از وب برای پیدا کردن اطلاعات گردشگری استفاده می‌کنند. اما با این وجود نزدیک به یک‌سوم از افراد از بیان نیازهای صریح گردشگری خود عاجز هستند (Zins, ۲۰۰۷). امروزه با افزایش میزان گردشگری و همچنین پیچیدگی اطلاعات این مشکل گسترده‌تر و پیچیده‌تر شده است و یکی از دلایل این پیچیدگی احساسی بودن فرآیند تصمیم‌گیری در صنعت گردشگری است. از سوی دیگر بنا به (Werthner et al., ۲۰۱۵) اگرچه وب دسترسی به اطلاعات قابل اطمینان را مهیا ساخته و در زمان و هزینه‌ها به‌طور چشمگیری صرفه‌جویی کرده و موجب راحتی در به دست آوردن اطلاعات در کسری کوتاهی از زمان شده است اما با این وجود طیف متنوعی از اطلاعات در وب سبب اطاعات بی‌شمار و افزونگی اطلاعات شده که این امر نیز منجر به مشکل در فرآیند تصمیم‌گیری و پیدا کردن اطلاعات منطبق با نیازهای انسان شده است. برای غلبه بر این مشکل وجود سامانه‌هایی که اطلاعات شخصی شده و مناسب را در اختیار کاربر قرار دهد به شدت احساس می‌شود. ریچی و همکاران سامانه‌های پیشنهاد گر را ابزارهای نرم‌افزاری و فن آوری‌هایی برمی‌شمارند که قابلیت ارائه اطلاعات مناسب به کاربر دارا است (Ricci et al., ۲۰۱۱). دودسته از سامانه‌های پیشنهاد گر توسط (Garcia et al., ۲۰۱۱) در گردشگری شناخته شده‌اند. ۱) سامانه‌های که مقصد گردشگری را معرفی می‌کنند (۲) سامانه‌های که در یک مقصد خاص مجموعه‌ای از فعالیت‌ها و مکان‌های گردشگری را معرفی می‌کنند. که در این پژوهش به نوع دوم خواهیم پرداخت. چندین رویکرد برای ساخت یک سامانه پیشنهاد گر هوشمند وجود دارد که به شرح زیر است:

۱. رویکرد مبتنی بر محتوا: توصیه‌ها مبتنی بر اقلامی است که در گذشته کاربر به آن‌ها علاقه نشان داده است.

۲. رویکرد پالایش مشارکتی: اقلامی توصیه می‌شود که توسط سایر کاربران پسندیده شده است.

۳. رویکرد مبتنی بر اطلاعات دموگرافیک<sup>۱</sup>: توصیه‌ها مبتنی بر داده‌های دموگرافیکی مانند سن و جنسیت است.

۴. رویکردهای ترکیبی یا به اصلاح هیبریدی: که ترکیبی از رویکردهای بیان شده است. (Neidhardt et al., ۲۰۱۵)

همچنین عمده ابزارها و فن‌های مورد استفاده از هوش مصنوعی شامل: سامانه‌های چندعاملی، فن‌های بهینه‌سازی، خوشه‌بندی‌های خودکار، مدیریت عدم قطعیت و نمایش دانش است که مورد آخری،

---

<sup>۱</sup> demographic

متداول‌ترین روش آن آنتولوژی<sup>۲</sup> است که برای بیان روشن یک فرم مفهومی از یک دامنه، که اجزای آن دسته‌هایی است که در یک شکل درختواره یا غیر درختواره مرتب می‌شوند (Borràs et al., ۲۰۱۴). مورد دیگر که در سامانه‌های پیشنهاد گر باید به آن پرداخت پروفایل<sup>۳</sup> کاربران است که اطلاعات مربوط به کاربران را در خود برای ارائه پیشنهادها به کاربر ذخیره می‌کند. انواع مختلفی از مدل‌های پروفایل کاربری توسعه یافته است. ساده‌ترین مدل، هر کاربر بر اساس گروهی از واژگان یا طبقه‌بندی‌ها برای ارتباط اقلام از پیش دسته‌بندی شده ارتباط پیدا می‌کند. این مدل‌های ساده می‌تواند با کمک ابزارها و رویکردهای هوش مصنوعی گسترده‌تر و پیچیده‌تر شود.

پیشنهادگرها با تمام مزایا، دارای مشکلاتی نیز هستند. با توجه به بررسی پژوهش‌های گذشته مدل‌های سامانه‌های پیشنهاد گر عمدتاً بنا بر اندازه‌گیری نزدیکی نمایه کاربر با مقصد، که هرکدام در یک بردار نمایش داده می‌شود که هر از یک مقادیر داخل این بردار نمایش‌دهنده مقدار عددی یک ویژگی است. مشکل این ایده نیاز به قضاوت دائمی خبرگان برای امتیازدهی به ویژگی‌های مقاصد است که روشی غیرمطمئن و در ابعاد بالای مقاصد گرد شگرتی سخت و زمان‌بر است. در پیشنهادگرها گرد شگرتی که از رویکرد مبتنی بر مشارکت استفاده شده است به‌طور عمده از فن‌ها و روش‌های خوشه‌بندی و دسته‌بندی‌های قواعد انجمنی استفاده می‌شود. ممکن است مجموعه‌ای از داده‌ها به‌طور مناسب قابل خوشه‌بندی نباشد و یا تعداد خوشه‌ها بسیار زیاد و یا بسیار کم که ارضا کننده الگوریتم پیشنهادگر نباشد. به‌طور معمول نمونه‌های موجود در هرکدام از خوشه‌ها، شباهت بسیاری باهم دارند و اگر صرفاً با توجه به یک خوشه پیشنهادت ارائه شود، ممکن است این پیشنهادت از تنوع مناسب برخوردار نباشد که این مشکل در ادبیات شخصی سازی بیش‌از حد یا به عبارتی دیگر به تکرار گرایی موسوم است. در پیشنهادگرهایی که از چندین خوشه برای پیشنهاد به کاربر استفاده می‌کنند، به‌طور معمول با توجه به خوشه‌ها اقلام پرترفدار هر خوشه یا متناظر با خوشه‌ها را انتخاب و به کاربر پیشنهاد می‌کنند، در این بین ممکن است اقلامی موجود باشند که در بین هیچ‌کدام از خوشه‌ها جزء اقلام پرترفدار نباشد ولی افراد به خصوصی به آن اقلام علاقه داشته باشند که در ادبیات به کنایه به چالش گوسفند خاکستری مشهور است. استفاده از روش‌های قواعد انجمنی نیز، زمانی که تعداد متغیرها بالا باشد و همچنین با تعداد داده‌های کم نمی‌تواند روشی اثربخش باشد به دلیل اینکه ممکن است قواعد استخراجی با توجه به بالا بودن تعداد متغیرها بسیار زیاد و باعث پیچیده‌تر شدن مسئله شود. در پیشنهادگرهای مبتنی بر مشارکت، شروع سرد مشکل رایجی است که برای حل این مشکل به‌طور معمول به سراغ دریافت اطلاعات صریح از کاربر می‌روند که معمولاً اطلاعات دموگرافیک از کاربران دریافت می‌شود که ممکن است برای کاربر ناخوشایند باشد و یا تعداد

---

<sup>۲</sup> ontology

<sup>۳</sup> profile



سوالات و یا عملگرها برای دریافت کاربر زیاد باشد که موجب کاهش رضایت و عدم همکاری کاربر شود. برای ارائه دادن اطلاعات پیشنهادگرهای مبتنی بر آنتولوژی بخش طراحی آنتولوژی می تواند چالش برانگیز و پرهزینه باشد. به دلیل این که عمدتاً ارتباط سطح های مختلف آنتولوژی توسط افراد خبره در حوزه های که پیشنهادگر برای آن توسعه پیدا می کند، تعیین می شود. بنابراین به ارائه روشی نیاز است که بتواند چالش های ارائه شده را رفع کند و پیشنهادات مناسب به کاربر ارائه کند که در نتیجه روش ارائه شده در این پژوهش مزایای زیر را خواهد داشت:

۱. موجب ارتقا فرآیند تصمیم گیری برای گردشگر می شود و گردشگر با اطلاعات مناسب و به جا تصمیم گیری می کند.
۲. گردشگر نیاز نیست تا اطلاعات و موارد بسیار زیادی را جستجو کند، این امر موجب صرفه جویی در زمان برای کاربر شده و همچنین از ارائه اطلاعات مازاد به سیستم های جستجوگر جلوگیری می کند که خود منجر به حفظ حریم اطلاعاتی کاربر می شود.
۳. در قیاس با سایر پیشنهادگرها که یک معیار ارزیابی روشن و دقیق ندارند، نظر به این که از روش های یادگیری ماشین و رویکرد دسته بندی استفاده خواهد شد این سامانه دارای معیار متریک برای اندازه گیری صحت پیشنهادات خواهد داشت و همچنین از روش های دیگر ارزیابی مانند مطالعه کاربری نیز استفاده خواهد کرد.

### ۱,۳ سوالات پژوهش

۱. چه نوع و چه تعداد دسته های گردشگری بر مبنای فعالیت های گردشگری و تفریحی تعریف شده در استان تهران موجود است؟
۲. میزان اثر گذاری دسته ها بر انتخاب شدن هر فعالیت برای کاربر چگونه است؟
۳. نتایج ارزیابی کاربر در قیاس با ارزیابی برون خطی چگونه است؟

### ۱,۴ اهداف پژوهش

- با توجه به مطالب بیان شده و اهمیت سامانه های هوشمند پیشنهاد گر در امر یاری رسانی تصمیم گیری و فراهم آوری اطلاعات متناسب با هر کاربر این پژوهش اهداف زیر را دنبال می کند:
۱. دستیابی به یک دسته بندی از گردشگران بر مبنای فعالیت های گردشگری.
  ۲. دستیابی به یک آنتولوژی منطقی برای دسته بندی مقاصد گردشگری و ارتباط سطح گروه ها و مفهوم بدون دخالت انسانی.

۳. طراحی یک سامانه پیشنهادگر که منجر برطرف سازی افزونگی اطلاعات و ارائه اطلاعات گردشگری سفارشی سازی شده و متنوع متناسب با هر کاربر می شود.
۴. یاری رسانی کاربر در بیان نیازهای گردشگری فقط با پاسخ به تعداد کمی سوال غیر دموگرافیک و گرفتن تصمیم برای انتخاب مقصد.

## ۱.۵ جنبه های نوآوری پژوهش

بیشتر پژوهش ها برای طراحی آنتولوژی از خبرگان آن حوزه استفاده می کنند و این فرآیند ممکن است سلیقه ای باشد. ما برای اولین بار در پیشنهادگرهای گردشگری سطح دوم آنتولوژی گردشگری در این پژوهش را که سطح گروه ها خوانده می شود از روش ها تحلیل عاملی بدست خواهیم آورد. همچنین ارتباط بین سطوح آنتولوژی نیز به طور معمول توسط خبرگان یا طراحان سامانه صورت می گیرد که ما در این پژوهش به ارائه روشی با مدل های دسته بندی چندبرچسب خواهیم پرداخت که ارتباط سطح گروه های آنتولوژی و سطح مفهوم آنتولوژی را بدون قضاوت انسانی برقرار می کند. برخی از پیشنهادگرها که از گروه ها و مدل های گردشگری استفاده می کنند و با ارزیابی این که کاربر به کدام گروه تعلق دارد موارد و اقلام همان گروه را به کاربر پیشنهاد می دهند که این امر منجر به تنوع کم پیشنهادات و شبیه بهم بودن پیشنهادات به کاربر می شود جهت رفع این مشکل برای اولین بار از ترکیب آنتولوژی و مدل های دسته بندی چندبرچسب برای ارائه پیشنهاد به کاربر استفاده خواهیم کرد.

## ۱.۶ مواد و روش انجام تحقیق

برای پژوهش حاضر مورد مطالعاتی ما برای طراحی سامانه های پیشنهادگر شهر تهران با توجه به اهمیت ژئوپلیتیکی، تراکم جمعیت بالا، بالاتر بودن ضریب نفوذ اینترنت و دارا بودن مکان های گردشگری متنوع انتخاب شده است. داده های که جمع آوری خواهد شد از نقطه نظر زمانی تجربی خواهد بود و از نقطه نظر نحوه گردآوری مبتنی بر روش های مطالعاتی است. سپس با ارائه روشی به دسته بندی و ساخت یک آنتولوژی از مراکز گردشگری تهران خواهیم پرداخت. در مرحله بعد با کمک جمع آوری داده ها به کمک پرسشنامه سعی در ارتباط گروه های گردشگری جامعه مورد آزمایش و فعالیت های سطوح مفهوم آنتولوژی و به تبع آن ارائه یک مدل توصیفی مناسب از گردشگر داریم. سپس به طراحی معماری یک مدل ترکیبی مبتنی بر پالایش مشارکتی و مبتنی بر یادگیری ماشین خواهیم پرداخت.

## ۱,۷ ساختار پژوهش

در فصل دوم که در رابطه با ادبیات موضوع است به بررسی مفاهیم پیشنهادگرها، تعاریف و فن‌ها و رویکردهای مورد استفاده در پیشنهادگرها خواهیم پرداخت. سپس به بررسی یادگیری ماشینی، الگوریتم‌ها و نحوه ارزیابی مدل‌های یادگیری ماشین خواهیم پرداخت. سپس مروری بر پژوهش‌های پیشین و اهداف آن‌ها و نحوه ارزیابی آن‌ها خواهیم داشت. در فصل سوم که موسوم به مواد و روش‌ها است مراحل و روش پژوهش خود و معماری سامانه پیشنهادگر ارائه شده در پژوهش را تشریح خواهیم کرد. در فصل چهارم به پیاده سازی و اجرای قدم به قدم مراحل پژوهش و ارائه گزارشات، ارزیابی و نتایج آن خواهیم پرداخت و در فصل آخر به نتیجه گیری از پژوهش حاضر، محدودیت‌های پژوهش و پیشنهاد برای پژوهش‌های آتی خواهیم پرداخت.

## **فصل دوم**

### **مروری بر ادبیات پژوهش**

## ۲,۱ مقدمه

فناوری اطلاعات و آمیختگی آن با زندگی انسان‌ها فصل جدیدی در رفتار و تعاملات انسانی ایجاد کرده است. فناوری اطلاعات موجب افزایش سرعت انجام فرآیندها، کاهش هزینه‌ها و سهولت دسترسی به خدمات شده است. این مزایا دو سویه است، یعنی هم برای کاربران و هم ارائه‌دهندگان خدمات سودمند بوده است؛ اما در کنار این مزایا، موجب مشکلاتی نیز شده است. این روزها کاربران می‌توانند مقادیر انبوهی از اطلاعات را در صفحات وب بیابند؛ اما بعضی اوقات همین مسئله می‌تواند مشکلی پیچیده برای کاربر جهت پیدا کردن اطلاعات مناسب باشد به گونه‌ای که اطلاعات مناسب می‌تواند در دریای انبوهی از اطلاعات غرق شود. اطلاعاتی در رابطه با مقصد گردشگری و منابع مرتبط با آن مانند تفرجگاه‌ها، رستوران‌ها، موزه‌ها و یا رویدادها، به‌طور معمول توسط گردشگران برای برنامه‌ریزی سفرشان جستجو می‌شود؛ اما ارزیابی این لیست بلند از گزینه‌ها برای گردشگران جهت انتخاب مناسب‌ترین برنامه برای آن‌ها، بسیار پیچیده و زمان‌بر است؛ بنابراین غلبه بر این مشکل می‌تواند قدم بزرگی جهت صرفه‌جویی در زمان و هزینه گردشگر داشته باشد و همچنین رضایت بیشتری را نیز فراهم آورد (Grün et al., ۲۰۱۷).

در این بخش ابتدا به بررسی پیشنهادگرها و تعاریف اولیه در رابطه با آن‌ها خواهیم پرداخت، سپس به تشریح رویکردها و فن‌ها مورد استفاده در این سامانه‌ها خواهیم پرداخت، سپس نحوه ارزیابی پیشنهادگرها را بررسی خواهیم کرد و در ادامه مروری بر پژوهش‌های پیشین در حوزه پیشنهادگرهای گردشگری و همچنین نقش‌ها و کلیشه‌ها در گردشگری خصوصاً از این جهت که ما نیز در پژوهش از این کلیشه‌ها و نقش‌ها استفاده کرده‌ایم، خواهیم داشت و به بررسی چالش‌ها و مشکلات خواهیم پرداخت. در قسمت بعد به بررسی روش‌ها و فن‌های مورد استفاده در پژوهش خواهیم پرداخت. ابتدا یادگیری ماشین و رویکردهای آن را معرفی می‌کنیم سپس تحلیل عاملی را بررسی می‌کنیم و بعد از آن به الگوریتم‌ها و مدل‌های دسته‌بندها استفاده‌شده در پژوهش و نحوه ارزیابی آن‌ها به‌طور خلاصه اشاره می‌کنیم و در آخر نوع

دیگری از دسته‌بندها، موسوم به دسته‌بندهای چندبرچسب که نقش پررنگ در پژوهش ما دارد را معرفی می‌کنیم و مثال‌هایی از پیشنهادگرها که از این نوع دسته‌بندها استفاده کرده‌اند را بررسی می‌کنیم.

## ۲.۲ سامانه‌های اولیه جستجوی اطلاعات

در این قسمت، سه نوع اولیه از سامانه‌های اولیه جستجوی اطلاعات را معرفی خواهیم کرد که عبارتند از:

### ۲.۲.۱ موتورهای جستجو<sup>۱</sup>

این موتورها کامپیوترهایی هستند که اطلاعات موجود در وب را نمایه<sup>۲</sup> می‌کنند. جستجو بر اساس کلیدواژه‌ها یا درختواره‌های سلسله‌مراتبی بر اساس عنوان‌های متفاوت صورت می‌گیرد. نتایج جستجو همراه است با تعداد زیادی از آدرس‌های اینترنتی که دربردارنده عناوین مرتبط با کلیدواژه‌های جستجو شده است. به‌طورمعمول، حجم عظیمی از اطلاعات به‌دست‌آمده از جستجو، موردپسند کاربر نیست.

### ۲.۲.۲ سامانه‌های یاریگر<sup>۳</sup>

آن‌ها به کاربر نحوه استفاده از یک برنامه خاص را تشریح می‌کنند. به‌طور مثال میکروسافت، دارای سامانه‌ای یاریگر است که اطلاعات در رابطه با ویژگی‌های کاربر و اقدامات او، وضعیت برنامه و واژگانی که کاربر آن‌ها را جستجو می‌کند را جهت محاسبه احتمال کمک موردنیازی که ممکن است کاربر به آن نیاز داشته باشد را جمع‌آوری می‌کند. این سامانه‌ها به‌طورمعمول به صفحاتی پیوند دارند که اطلاعات بیشتری در اختیار قرار می‌دهند.

### ۲.۲.۳ سامانه‌های پالایش اطلاعات<sup>۴</sup> و سامانه‌های بازبایی<sup>۵</sup>

هدف این سامانه‌ها بدست آوردن مرتبط‌ترین اطلاعات به کاربر و به حداقل رساندن موارد غیر مرتبط است؛ بنابراین، سامانه‌های پالایشی مقادیر عظیمی از اطلاعات ناخواسته را حذف می‌کند. اگرچه زمانی مفید محسوب می‌شوند که علایق کاربر را با یک رویکرد خودکار در نظر بگیرند. اگر این سامانه‌ها با هوش مصنوعی همراه شوند، می‌توانند محتوای موردنظر کاربر را به‌جای مقدار عظیمی از داده‌های مرتبط و غیر مرتبط فراهم کنند (Ciurana Simó, ۲۰۱۲).

---

<sup>۱</sup> Search engines

<sup>۲</sup> index

<sup>۳</sup> Helping system

<sup>۴</sup> Information filtering system

<sup>۵</sup> Retrieval system



## ۲,۳ سامانه پیشنهادگر<sup>۶</sup>

این سامانه‌ها که زیرمجموعه‌ای از سامانه‌های تصمیم‌یار<sup>۷</sup> هستند، به‌عنوان ابزاری برای پیشنهاد اقلامی از داده‌ها بر مبنای ترجیحات کاربر استفاده می‌شود. این سامانه‌ها اطلاعات ارزشمندی را برای کاربر مهیا می‌کنند تا کاربر بتواند بر مبنای اولویت‌ها و دغدغه‌های خود به‌درستی تصمیم‌گیری کند. سامانه‌های پیشنهادگر به‌طور معمول روش‌شناسی خود را از سه زمینه بکار می‌گیرند که این سه زمینه، بازیابی اطلاعات<sup>۸</sup>، تعامل انسان کامپیوتر<sup>۹</sup> و داده‌کاوی<sup>۱۰</sup> است. این سامانه‌های نقش بسیار پررنگی در زندگی روزمره انسان‌ها بازی می‌کنند و بسیاری از وب‌سایت‌ها مانند آمازون، نتفلیکس، یاهو و ... از این سامانه در وب‌سایت‌ها و برنامه‌ها خود استفاده می‌کنند. پیشنهادگرها در حوزه وسیعی از زمینه‌ها بکار گرفته شده‌اند که یکی از مهم‌ترین و پرکاربردترین آن‌ها حوزه گردشگری است (Ciurana Simó, ۲۰۱۲). محور اصلی پیشنهادگرها شخصی‌سازی اطلاعات برای کاربر، بنا به ترجیحات، علائق و محدودیت‌های کاربر است. هدف پالایش گزینه‌های بی‌ربط و مهیا کردن اطلاعات مناسب و شخصی سازی شده برای کاربر است. در زمینه گردشگری، هدف پیشنهادگرها انطباق ویژگی‌ها و نیازهای گردشگر با مقاصد گردشگری و تفریحی است. این سامانه‌ها زمانی مفیدتر واقع می‌شوند که بتوانند به‌طور خودکار علائق کاربر را از اطلاعات صریح و یا غیر صریح کاربر استنباط کنند. داده‌های صریح به شیوه‌های متفاوتی می‌تواند جمع‌آوری شود، مثلاً زمانی که کاربر میزان علاقه خود را به فرهنگ و هنر از طریق یک فرم پرسشنامه‌ای مشخص می‌کند و همچنین اطلاعات غیر صریح از طریق تحلیل رفتار کاربر استنباط می‌شود (Neidhardt et al., ۲۰۱۵). در بسیاری از موارد پیشنهادگر نه تنها ترجیحات کاربر را در نظر می‌گیرد بلکه به‌طور پویا اطلاعات سفر گردشگر را نیز تحلیل می‌کند. این رویکرد به‌خصوص زمانی که گردشگر در مقصد موردنظر خود است و قصد دارد برنامه گردش خود را تنظیم کند، بسیار سودمند واقع می‌شود. این اطلاعات می‌تواند، موقعیت مکانی گردشگر، زمان بازدید مکان‌های گردشگری، وضعیت آب‌وهوا و ... باشد. به‌طور مثال بازدید از موزه در برنامه گنجانده شده است ولی به دلیل دور بودن یا بسته بودن موزه در زمان حاضر، برنامه بازدید موزه می‌تواند به روز و زمان دیگری موکول شود (Borràs et al., ۲۰۱۴).

---

<sup>۶</sup> Recommender system

<sup>۷</sup> Decision support systems

<sup>۸</sup> Information retrieval

<sup>۹</sup> Human Computer Interaction

<sup>۱۰</sup> Data mining

### ۲,۳,۱ تعاریف اولیه در سامانه‌های پیشنهادگر

لازم است برای درک مفهوم سامانه پیشنهادگر مفاهیم ابتدایی زیر را بررسی کنیم. در سامانه‌های پیشنهادگر به کاربری که پیشنهاد جاری در سیستم برای وی در حال پردازش و آماده شدن است کاربر فعال یا کاربر هدف گفته می‌شود.

الگوریتم‌های به‌کاررفته در این سامانه‌ها از ماتریس‌ای به نام ماتریس رتبه‌دهی یا نرخ دهی استفاده می‌کنند ساختار ماتریس رتبه‌ها بدین گونه است که در آن هر سطر ماتریس نمایانگر یک کاربر و هر ستون آن معرفی یک کلمه خاص است نمونه‌ای از این ماتریس در جدول ۱-۲ آمده است این جدول مثالی از ماتریس نرخ دهی کاربران برای پیشنهاد اقلام است. نرخ‌ها بین ۱ تا ۵ است نماد  $\emptyset$  به این معناست که کاربر نرخ به آن آیتم نداده است بنابراین سامانه پیشنهادگر بایستی بتواند اقلام و موارد نرخ داده نشده و انواع مناسب پیشنهاد بر اساس این نرخ‌ها را تخمین زده و پیش‌بینی نماید.

از عبارت مصرف کردن در سامانه‌های پیشنهادگر زمانی استفاده می‌کنند که کاربر پیشنهاد ارائه شده توسط سامانه را می‌پذیرد، به عبارت دیگر وقتی کاربری پیشنهادی را که توسط سامانه به او شده است را می‌پذیرد گفته می‌شود کاربر پیشنهاد را مصرف کرده است. این پذیرش می‌تواند به شکل‌های مختلفی باشد مثلاً کاربر کتاب پیشنهادی را می‌خرد، سایت پیشنهادی را مرور می‌کند و یا به شرکت خدماتی که به او پیشنهاد شده مراجعه می‌کند.

مورد بعدی مفهوم تابع سودمندی است. به منظور توضیح نحوه پیاده‌سازی سامانه‌های پیشنهادگر به لحاظ این که قصد داریم به کمک آن یک مدل کلی ریاضی از سامانه‌های پیشنهادی را ارائه دهیم با مفهوم تابع سودمندی آشنا می‌شویم. درواقع یک سامانه پیشنهادگر را می‌توان با این نگاشت همسان دانست و مدل کرد:

$$u: C * S \rightarrow R$$

رابطه ۱-۲

جدول ۱-۲ ماتریس نرخ دهی کاربران برای پیشنهاد اقلام (آذین، خداپرست، ۱۳۹۱).

قلم A	قلم B	قلم C	قلم D
$\emptyset$	۴	۵	۳
۵	۲	$\emptyset$	۱
۲	۴	۵	$\emptyset$

کاربر A

کاربر B

کاربر C

فرض کنید  $C$  تمامی کاربران و  $S$  مجموعه اقلام در دسترس باشند. تابعی را که میزان مفید و مناسب بودن کالای  $s \in S$  را برای کاربر  $c \in C$  را محاسبه می‌کند با  $u$  نمایش می‌دهیم، که در آن  $R$ ، مجموعه‌ای کاملاً مرتب بر اساس میزان اهمیت است.  $R_{cs}$  نرخ‌گذاری و رتبه‌ای است که کاربر  $c$  بر روی مورد  $s$  انجام می‌دهد. هر کدام از عناصر  $S$  را می‌توان با مجموعه‌ای از خصوصیات خود (نمایه)، توصیف نمود. برای مثال، محصولی مثل فیلم را می‌توان با مشخصه‌هایی چون عنوان فیلم، کارگردان، طول زمان فیلم، تاریخ تولید و غیره ثبت کرد. همچنین عناصر مجموعه  $C$  را نیز می‌توان بر اساس ویژگی‌های مثل سن، جنسیت و غیره ثبت کرد. سپس برای هر کاربر  $c \in C$  کالایی مانند  $s \in S$  را که سودمندی کاربر را بیشینه نماید، به‌طور قراردادی به‌صورت زیر تعریف نماییم:

$$\forall c \in C, S_c = \operatorname{argmax}_{s \in S} u(c, s) \quad \text{رابطه ۲-۲}$$

مورد آخر، فرآیند شناخت و طبقه‌بندی کاربران است. باید توجه داشت اطلاعاتی که یک سامانه پیشنهادگر از کاربران به دست می‌آورد هم از لحاظ نوع و هم از لحاظ نحوه‌ی به دست آمدن متفاوت هستند. به دریافت اطلاعات توسط سامانه و به‌طور کلی فرآیند شناخت کاربران نرخ‌گذاری گفته می‌شود. در سامانه‌های پیشنهادگر سودمندی یک پیشنهاد معمولاً به‌وسیله نرخ‌گذاری مشخص می‌شود و بیان می‌کند چگونه کاربر خاصی، محصولی را دوست دارد.

به‌طور کلی فرآیند شناخت کاربران به دو روش زیر انجام می‌شود:

روش مبتنی بر دانش (صریح) این روش مدل‌های ثابتی از کاربران ایجاد کرده و سپس کاربران در نزدیک‌ترین گروه‌های مشابه و هم سلیقه گروه‌بندی می‌شوند. اطلاعات دانش در مورد کاربران از راه‌های مختلفی به دست می‌آید. اولین راه پرسش از کاربر برای انتخاب بین انواع مفاهیم و خدمات است. علیرغم مزیت کسب اطلاعات مستقیم و دقیق از کاربر، این روش همیشه و برای همه کاربران پاسخگو نیست. در مواردی که کاربر حوصله پاسخگویی به سؤالات را ندارد این روش ممکن است که کاملاً اطلاعات اشتباهی را ارائه نموده و گاهی مجبور به استناد به اطلاعات قدیمی و منسوخ می‌شود لذا ممکن است بهتر باشد این اطلاعات به‌خصوص اطلاعات غیرثابت و متغیر از کاربر به‌طور غیرمستقیم گرفته شود و به‌صورت مفهومی پردازش و سپس بعد از آن به‌عنوان پایه‌ای برای پیشنهاد استفاده شود.

روش مبتنی بر رفتار (ضمنی) این روش، از مدل رفتار کاربر استفاده می‌نماید در این روش با استفاده از فناوری‌های یادگیری بر اساس فعالیت‌های کاربر در وب سایت‌ها رفتارهای باقاعده او را شناسایی و سپس یک مدل به نام مدل کاربر ساخته می‌شود. مدل کاربر مجموعه‌ای از اطلاعات درباره رفتار کاربر است و به شناسایی بیشتر کاربر که چه کرده و حدس اینکه در آینده چه خواهد کرد کمک می‌نماید. به‌طور نمونه در بعضی از سامانه‌های پیشنهادگر ماشین‌های یادگیری تعداد خریده‌ها و جستجوهای هر کاربر را ذخیره و سپس بررسی نموده و علایق او را بر این اساس استخراج می‌نمایند.

روش ترکیبی برخی از سامانه‌های پیشنهادگر، از هر دو روش برای تکمیل فرآیند شناسایی کاربران خود و سپس ارائه پیشنهاد بهره می‌برند. از جمله این سامانه‌ها، می‌توان از <sup>۱۱</sup>UPE نام برد. UPE یک سامانه پیشنهادگر است که پیشنهادهای شخصی در مورد صفحاتی که دارای کاتالوگ‌های جذاب برای کاربران است، ارائه می‌دهد. در این سامانه خصوصیت کاربران هم از نوع متغیر و هم از نوع ثابت است. خصوصیات ثابت اطلاعاتی هستند که برای هر کاربر مشخص بوده و تغییر نمی‌کند و یا به ندرت تغییر می‌کند و از طریق برگه‌های ثبت نام و اطلاعات درخواستی از کاربر دریافت می‌شود (مانند نام کاربر). اطلاعات متغیر یک کاربر خصوصیتی از کاربر است که ممکن است در هر بازدید تغییر نماید (مانند اولویت‌های کاربر) (آذین، خداپرست، ۱۳۹۱).

در جدول ۲-۲ نمونه‌ای از انواع داده‌ها و اطلاعاتی که به کمک روش‌های فوق به دست می‌آیند نشان داده شده است.

جدول ۲-۲ نمونه‌هایی از روش‌های جمع‌آوری اطلاعات و شناخت کاربران (آذین، خداپرست، ۱۳۹۱).

روش	نحوه کسب اطلاعات	نمونه آیتم‌ها
مبتنی بر دانش	داده‌های شخصی	نام، جنس، شغل، درآمد و آدرس
	پرس‌وجوهای شخصی	سطح تخصص کاربر، حوزه‌های موردعلاقه کاربر و علایق مرتبط
مبتنی بر رفتار	جنبه‌های آماری	تعداد بازدیدها، میزان زمان صرف شده در هر بازدید صفحه، ترتیب URL های بازدید شده و فرآیند جستجو
ترکیبی	فن‌های مفهومی در ترکیب داده‌های صریح و ضمنی	اعتماد میان کاربران، بازدیدهای مشابه، میزان تخصص در حوزه خدمات، خریدهای محتمل و غیره

## ۲,۳,۲ فرآیند شخصی‌سازی اطلاعات<sup>۱۲</sup>

نتیجه فرآیند یک سامانه پیشنهادگر، ارائه مجموعه‌ای از اقلام و عناصر در یک رابط کاربری است. فرآیند شخصی‌سازی به سه مرحله زیر تقسیم می‌شود:

<sup>۱۱</sup> User profile engine

<sup>۱۲</sup> Personalizing process

جمع‌آوری اطلاعات کاربر: اطلاعات کاربر جهت شناخت بهتر از او جمع‌آوری می‌شود. این اطلاعات در نمایه کاربر ذخیره می‌شود.

پیشنهاد موارد: بر اساس دانشی که سامانه از کاربر بدست آورده است، مواردی را به کاربر پیشنهاد می‌کند. درجه رضایت: اندازه‌گیری تأثیرگذاری شخصی‌سازی اطلاعات بنا به نظرات کاربر نسبت به موارد پیشنهاد شده است. این مرحله به‌عنوان یک بازخورد از مراحل قبل استفاده می‌شود و از آنجایی که نظرات و عقاید کاربر را نیز دخیل می‌کند بنابراین شناخت سامانه از کاربر بهتر و بیشتر می‌شود (Ciurana Simó, ۲۰۱۲).

### ۲,۳,۳ مزایا و معایب پیشنهادگرها

پیشنهادگرها از جهات گوناگونی هم برای کاربر و هم برای صاحبان کسب‌وکار سودمند هستند. برخی از مزیت‌های آن‌ها به شرح زیر است:

- گنجاندن پیشنهادگرها در صفحات وب و برنامه‌های کامپیوتری، امکان سفارشی‌سازی خدمات و اطلاعات را برای کاربر با توجه به نیازها و علاقه‌مندی‌هایش فراهم کرده است.
- مزیت اصلی پیشنهادگرها زدودن اطلاعاتی است که کاربر به آن‌ها نیازی ندارد.
- مزیت دیگر افزایش بازدید و استفاده از وب‌سایت و برنامه‌های دارای پیشنهادگر می‌شود.
- پیشنهادگرها می‌توانند بازدیدکنندگان را تبدیل به مشتریان دائمی کنند.
- زمانی که اطلاعات با توجه به نیازها و علائق کاربر نمایش داده می‌شود بنابراین امکان انتخاب و یا خرید اقلام بیشتر می‌شود بدین ترتیب موجب افزایش اقلام سفارشی یا موارد انتخابی کاربر می‌شود.
- کاربران با حداقل اطلاعات می‌توانند انتخاب‌های مناسب‌تری داشته باشند.
- همچنین بازخوردهای کاربران به پیشنهادها این سامانه‌ها می‌تواند اطلاعات ارزشمندی را برای کسب‌وکارها جهت ساماندهی و برنامه‌ریزی برای محصولات و خدمات، فراهم کند.
- اما پیشنهادگرها می‌توانند دارای معایبی نیز باشند و در سطح اجتماعی مشکل‌ساز شوند. به‌طور مثال:
- هدایت کاربران به سمت کالا و موارد پرسودتر
- هدایت به سمت موارد خاص و موردنظر صاحبان وب‌سایت و برنامه‌ها
- مهم‌تر از همه جنبه امنیتی پیشنهادگرها است. یکی از ویژگی‌های برجسته پیشنهادگرها شناسایی و درک نیازهای کاربر است که همین امر می‌تواند موجب دردسرسازی برای کاربر شود.

اگرچه که پژوهش‌ها و نوآوری‌هایی در این زمینه انجام شده است ولی با این وجود همچنان کافی به نظر نمی‌رسد (Shoval, ۲۰۰۰).

#### ۲,۳,۴ رویکردها و فن‌های مورد استفاده در پیشنهادگرها

شافر بیان می‌دارد که پیشنهادگرها می‌توانند بر اساس درجه شخصی سازی اطلاعات برای کاربران که شامل دقت و مفید بودن پیشنهادها است دسته‌بندی شوند. درجه شخصی سازی اطلاعات می‌تواند از کم به زیاد، بدون شخصی سازی، شخصی سازی شده کوتاه مدت و شخصی سازی شده باثبات و بلندمدت باشد. پیشنهادگرهای بدون شخصی سازی شده<sup>۱۳</sup>، پیشنهادگرهایی هستند که بدون در نظر گرفتن ویژگی‌ها و علائق کاربر و فقط با توجه به انتخاب‌های پربازدید از اقلام یا موارد، آن‌ها را به کاربر پیشنهاد می‌کند. البته پژوهش‌ها در حوزه پیشنهادگرها به دلیل محدودیت این دسته از پیشنهادگرها (شخصی سازی نشده) تمرکز ندارد (Ricci, ۲۰۱۱).

از سوی دیگر پیشنهادگرهایی که هدف آن‌ها شخصی سازی اطلاعات برای کاربر هستند، نظر به دریافت اطلاعات از کاربر و ایجاد شناخت از او بسیار پیشرفته‌تر و قدرتمندتر هستند. بدین ترتیب که هر کاربری موارد پیشنهادی متفاوت از دیگری را با توجه به ویژگی‌ها، علائق و محدودیت‌های خود می‌بیند. به طور مثال تریپ ادوایزر<sup>۱۴</sup> پیشنهادهایی را بر اساس ویژگی‌هایی دموگرافیک و اجتماعی فرد ارائه می‌کند. پیشنهادگرها به طور عمومی و سنتی بر اساس نحوه تحلیل و پالایش اطلاعات به دسته‌های مبتنی بر محتوا<sup>۱۵</sup>، مبتنی بر پالایش مشارکتی<sup>۱۶</sup>، مبتنی بر دانش<sup>۱۷</sup> و روش‌های ترکیبی<sup>۱۸</sup> تقسیم می‌شوند. شکل ۱-۲ نمایی از دسته‌بندی این رویکردها را نمایش می‌دهد. علاوه بر روش‌ها و رویکردهای اشاره شده، روش‌ها و فن‌های جدیدتر نیز ابداع و توسعه داده شده‌اند. در بخش اول به توضیح رویکردهای عمومی و در ادامه به تشریح رویکردهای جدیدتر برای پیشنهادگرها خواهیم پرداخت.

<sup>۱۳</sup> None-personalized

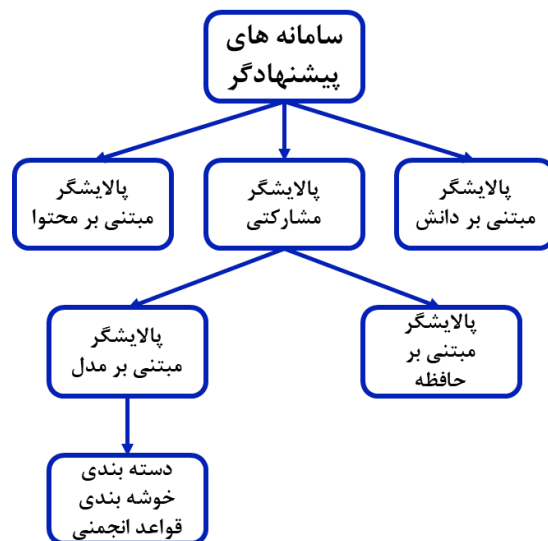
<sup>۱۴</sup> Trip advisor (www.tripadvisor.com)

<sup>۱۵</sup> Content based filtering

<sup>۱۶</sup> Collaborative based filtering

<sup>۱۷</sup> Knowledge based

<sup>۱۸</sup> Hybrid system



شکل ۲-۱ دسته بندی رویکردی پیشنهادگرها (Jannach et al., ۲۰۱۰)

پالایشگر مبتنی بر محتوا بنا به مقاله (Lu et al., ۲۰۱۵) در این رویکرد سامانه‌ها درجه شباهت بین کاربر و مورد پیشنهادی را اندازه‌گیری می‌کند. این فرآیند با مقایسه ویژگی‌های محصول نظر به ترجیحات کاربر انجام می‌شود. بنابراین فرض می‌شود که هم کاربر و هم موارد منتخب برای پیشنهاد دارای ویژگی‌های مشترک هستند. خروجی فرآیند مقایسه معمولاً امتیاز کلی عملکرد را نمایش می‌دهد، که این امتیاز بیانگر درجه انطباق نمایه کاربر با موارد منتخب برای پیشنهاد است. هر کدام از موارد منتخب امتیاز بیشتری کسب کنند یعنی عملکرد بهتری دارند و میزان انطباق آن‌ها بالا است. بعضی اوقات این رویکرد به تاریخچه امتیازدهی کاربر نیز می‌پردازد. در این رویکرد سامانه باید شناخت و دانش دقیقی از کاربر جهت ارائه پیشنهاد داشته باشد. اما یکی از مشکلات این رویکرد شروع سرد<sup>۱۹</sup> است. از این جهت که، زمانی کاربر تازه‌وارد سامانه می‌شود هیچ‌گونه اطلاعات تاریخچه‌ای از او وجود ندارد. قاعده کلی این نوع از سامانه‌ها عبارت‌اند از:

- تحلیل مشخصه‌های مواردی که توسط یک کاربر به‌خصوص ترجیح داده شده جهت پیدا کردن مشخصه‌های مشترکی که باعث تشخیص و تمیز دادن این موارد می‌شود. این مشخصه‌ها در نمایه کاربر ذخیره می‌شود.
  - مقایسه هر کدام از موارد با نمایه کاربر، بنابراین مواردی که به نمایه کاربر نزدیک‌تر و شبیه‌تر باشند برگزیده و پیشنهاد می‌شوند.
- دو رویکرد کلی و متداول برای رتبه‌دهی و محاسبه فاصله وجود دارد:

<sup>۱۹</sup> Cold start

۱. زمانی که نمایه کاربر با برداری از رتبه‌دهی‌ها از موارد موردعلاقه کاربر موجود است. هر کدام از رتبه‌ها می‌تواند به عنوان امتیاز عملکرد، جهت ارزیابی گزینه‌های پیشنهادی باشد؛ بنابراین هدف نهایی محاسبه امتیاز علاقه‌مندی فرد به یک گزینه محتمل برای پیشنهاد است. رویکرد ساده می‌تواند استفاده از یک عملگر تجميع کننده برای آمیختن نرخ‌های رتبه دهی کاربر به مفاهیم که سازنده یک گزینه پیشنهادی محتمل است، باشد. (Huang and Bian, ۲۰۰۹)

۲. زمانی که موارد یا اقلام و کاربر با مجموعه‌ای از کلیدواژه‌ها تعریف و توصیف شده‌اند، شباهت سنجی می‌تواند اعمال شود. به طور مثال می‌توان از آنتولوژی<sup>۲۰</sup> برای توصیف کاربر و اقلام استفاده کرد و از فاصله کسینوسی برای محاسبه میزان شباهت کاربر و اقلام استفاده کرد.

بنابراین به طور کلی می‌توان از روش‌های سنتی بازایی اطلاعات مانند شباهت سنجی بر اساس فاصله کسینوسی استفاده کرد و یا استفاده از روش‌های یادگیری آماری<sup>۲۱</sup> و یادگیری ماشین<sup>۲۲</sup> که می‌توانند از داده‌های قبلی قواعد و الگوهایی را جهت شناخت کاربر استخراج کنند استفاده کرد (Lu et al., ۲۰۱۵). مشکل اصلی در این نوع سامانه‌های پیشنهادگر، خصوصی سازی شده بیش از حد اطلاعات است که موجب می‌شود گزینه‌های پیشنهادی بسیار شبیه هم باشند و به طور کامل پوشش دهنده علائق کاربر نباشد.

**پالایه شگر مبتنی بر مشارکت:** سامانه پیشنهادگر مبتنی بر این روش، به افراد کمک می‌کند تا به کمک نظرات افرادی که به آن‌ها از لحاظ عقیده شبیه هستند انتخاب خود را انجام دهند. شباهت بین کاربران بر اساس امتیازی که آن‌ها به اقلام موارد در لیست داده‌اند محاسبه می‌شود. زمانی که سامانه دریابد که چه افرادی بر اساس علائق و انتخاب‌ها به هم نزدیک هستند، موارد پسند شده توسط کاربران دیگر به کاربر موردنظر پیشنهاد می‌شود. در این رویکرد بازخورد، جهت اینکه دریابیم کدام مورد پیشنهادی پسند واقع شده و کدام پسند نشده است، ضروری و لازم است. دو رویکرد کلی در این دست از سامانه‌های پیشنهادگر موجود است. رویکرد اول مبتنی بر کاربر یا روش مبتنی بر مدل گفته می‌شود و رویکرد دوم مبتنی بر اقلام یا موارد یا رویکرد مبتنی بر حافظه است. در رویکرد اول، کاربر موارد و اقلام را دریافت می‌کند که توسط یار کاربران شبیه به او موردپسند واقع شده است، به طور مثال برای ساخت مدل از کاربر از الگوریتم دسته‌بند مانند بیز استفاده می‌کنند؛ اما در رویکرد دوم، کاربر پیشنهاد‌های از اقلام یا موارد را

---

<sup>۲۰</sup> ontology

<sup>۲۱</sup> Statistical learning

<sup>۲۲</sup> Machine learning



دریافت می‌کند که شبیه به اقلام یا موارد موردپسند شده کاربر درگذشته است. این شباهت می‌تواند به وسیله همبستگی پیرسون<sup>۲۳</sup> و یا فاصله کسینوسی اندازه‌گیری شود. دو ایراد عمده به این نوع از سامانه‌ها وارد است: شروع سرد که قبلاً هم اشاره شد و دیگری تنک<sup>۲۴</sup> بودن داده‌ها است، یعنی به گونه‌ای که اقلام و مواردی که کاربر به آن‌ها امتیاز داده در مقابل کل اقلام و موارد بسیار کم باشد. البته مورد دیگری هم این است که برخی از کاربران شباهت کمی به کاربران دیگر دارند و به عبارتی پیدا کردن نقاط مشترک با کاربران دیگر مشکل و یا غیرممکن است (Thiengburanathum, ۲۰۱۸).

پالایشگر مبتنی بر دانش این دسته از سامانه‌های پیشنهادگر، موارد پیشنهادی آن‌ها بر اساس شناخت کاربر و اقلام و یا بر اساس رابطه آن‌ها، است. پیشنهادها برای کاربر بر اساس استنباط این است که چگونه موارد یا اقلام، یا توجه به نیاز کاربر برای او مناسب است. آنتولوژی یک روش نمایش رسمی از دانش است که مفاهیم دامنه و ارتباط این مفاهیم را نمایش می‌دهد. آنتولوژی برای نمایش دامنه دانش در سامانه‌های پیشنهادگر مبتنی بر دانش استفاده می‌شود. شباهت منطقی بین موارد و اقلام می‌تواند بر اساس آنتولوژی دامنه محاسبه شود (Borràs et al., ۲۰۱۴).

رویکرد ترکیبی برای رسیدن به عملکرد بهتر و غلبه بر ضعف‌های روش‌های ذکرشده، پیشنهادگرهای ترکیبی که دو یا چند از بهترین ویژگی‌های سایر روش‌ها را باهم ادغام می‌کنند پا به میدان گذاشتند. بنا بر بروک و همکاران، هفت روش برای ترکیب فن‌ها و روش‌هایی قبلی که اشاره شد وجود دارد که عبارتند از: وزن‌دار کردن<sup>۲۵</sup>، آمیختن<sup>۲۶</sup>، تعویض کردن<sup>۲۷</sup>، ترکیب ویژگی‌ها<sup>۲۸</sup>، افزایش ویژگی‌ها<sup>۲۹</sup>، آبشاری<sup>۳۰</sup> و فرا-مرحله<sup>۳۱</sup> است. یکی از مرسوم‌ترین ترکیب‌ها، ترکیب پالایشگر اشتراکی با سایر روش‌ها برای غلبه بر مشکل شروع سرد و تنک بودن داده‌ها است (Borràs et al., ۲۰۱۴).

---

<sup>۲۳</sup> Pearson correlation

<sup>۲۴</sup> sparse

<sup>۲۵</sup>Weighted

<sup>۲۶</sup> Mixture

<sup>۲۷</sup> Switching

<sup>۲۸</sup> Feature combination

<sup>۲۹</sup> Feature augmentation

<sup>۳۰</sup> Cascade

<sup>۳۱</sup> Meta-level

روش‌های مبتنی بر هوش محاسباتی<sup>۳۲</sup> رویکرد هوش محاسباتی شامل روش‌های بیز، شبکه‌های عصبی مصنوعی، خوشه‌بندی، الگوریتم‌های ژنتیک و مجموعه‌های فازی است. در سامانه‌های پیشنهادگر این روش‌ها استفاده بسیار گسترده‌ای دارند.

یک دسته‌بند بیز یک روش احتمالی برای حل مسائل دسته‌بندی است. استفاده از دسته‌بند بیز برای روش‌های مبتنی بر مدل بسیار پر استفاده بوده و در سامانه‌های مبتنی بر محتوا اعمال می‌شوند. زمانی که یک شبکه بیز در یک سامانه پیشنهادگر به کار گرفته می‌شود، هر گره در شبکه بیانگر یک قلم یا مورد است.

شبکه‌های عصبی مصنوعی مجموعه‌ای از گره‌های پیونددار است که خود این پیوندها دارای وزن هستند. این شبکه‌ها با الهام از مغز انسان ساخته و طراحی شده‌اند و برای سامانه‌های مبتنی بر مدل استفاده می‌شوند (Amatriain et al., ۲۰۱۱).

لو و همکاران (Lu et al., ۲۰۱۵) به نقل از (Amatriain et al., ۲۰۱۱) بیان می‌دارد خوشه‌بندی روشی است که داده‌ها (می‌تواند شامل اقلام یا موارد در سامانه‌های پیشنهادگر باشد) بر اساس شباهتشان در گروه‌های دسته‌بندی می‌شوند. یکی از کاربردهای دیگر خوشه‌بندی خلاصه سازی و هموار سازی داده‌ها برای بهتر شدن و کاهش محاسبات برای پردازش‌های ثانویه، مانند نزدیک‌ترین همسایه<sup>۳۳</sup> است. به طور مثال مقاله (Xue et al., ۲۰۰۵) از روش خوشه‌بندی برای ایجاد گروه‌هایی از کاربران و ارائه پیشنهاد موارد مورد پسند کاربران آن گروه به کاربر خاصی که حدس زده می‌شود به آن گروه نزدیکی و شباهت دارد، استفاده کرده است.

همچنین به نقل از (Kim and Ahn, ۲۰۰۸) الگوریتم‌های ژنتیک، یک روش جستجوی تصادفی است که برای بهینه‌سازی متغیرها در مسائل که دارای تابع هدف هستند مناسب است. این الگوریتم‌ها به طور عمده در دو جنبه از پیشنهادگرها استفاده می‌شوند که شامل خوشه‌بندی و مدل‌های ترکیبی از کاربر است. الگوریتم ژنتیک به همراه الگوریتم خوشه‌بند کا-میانگین<sup>۳۴</sup> به مجموعه داده‌های واقعی برخط در یک فروشگاه زنجیره‌ای جهت ایجاد گروه‌های مناسب برای ارائه پیشنهاد استفاده شد. نتیجه آن بهبود در تفکیک گروه‌ها بندی‌ها بود. همچنین از این الگوریتم‌ها برای بهینه‌سازی تابع شباهت سنجی، استفاده شد که نتیجه آن نمایانگر بهبود عملکرد و افزایش سرعت محاسبات شد (Lu et al., ۲۰۱۵).

لو و همکاران (Lu et al., ۲۰۱۵) به نقل از (Zenebe and Norcio, ۲۰۰۹) بیان می‌دارد، مجموعه‌های فازی، ارائه‌دهنده مجموعه‌ای از روش‌ها برای مدیریت عدم قطعیت غیر تصادفی است. برای مدیریت و اداره

<sup>۳۲</sup> Computational intelligence-based recommendation techniques

<sup>۳۳</sup> K-nearest neighbor

<sup>۳۴</sup> K-means

کردن داده‌هایی که دقیق و شفاف نیستند و دسته‌هایی که به‌طور دقیق تفکیک نشده‌اند بسیار مناسب است. به‌طور مثال در مسیریابی، در این رویکرد سعی در برآورد عدم قطعیت‌ها و تصمیمات کاربر دارد تا بتواند پیشنهاد‌های هوشمندانه‌تری ارائه دهد.

پیشنهادگرهای مبتنی بر شبکه‌های اجتماعی<sup>۳۵</sup> استفاده از تحلیل شبکه‌های اجتماعی در پیشنهادگرها نتیجه رشد چشمگیر ابزارهای شبکه‌های اجتماعی مبتنی بر وب در سال‌های اخیر است. برای کمک به تجربه کاربری، سامانه‌های پیشنهادگر به‌طور فزاینده‌ای کاربر را در تعاملات اجتماعی با دیگر کاربران، مانند خریدهای برخط، ارائه نظرات اجتماعی، برچسب‌گذاری‌های اجتماعی و ... قرار می‌دهند. این روند موجب یک فرصت برای پیشنهادگرها می‌شود تا از وابستگی کاربران به شبکه‌های اجتماعی برای غلبه بر مشکلاتی مانند تنگ بودن داده‌ها که در رویکردهای پالایشگر اشتراکی وجود دارد، استفاده کنند. اعتماد یکی از بحث‌برانگیزترین موارد در رابطه با شبکه‌های اجتماعی است. تصور کنید شخصی قصد خرید چیزی را داشته باشد، در این حالت نظرات دوستان در خرید او بسیار تأثیرگذارتر از تبلیغات اینترنتی است؛ بنابراین شبکه‌های اجتماعی می‌تواند یک منبع مهم و مناسب برای ارائه پیشنهاد باشد. در شبکه‌های اجتماعی واژه اعتماد به این‌گونه تعریف می‌شود که چگونه یک شخص به شخص دیگری در رابطه با خرید و نظر او در مورد چیزی اعتماد دارد. این مورد به اثبات رسیده که بین اعتماد کاربران و شباهت و اشتراکاتشان همبستگی وجود دارد. محققان مجموعه‌ای از مطالعات جهت ادغام اعتماد در سامانه‌های پیشنهادگر انجام داده‌اند. مقدار برآورد اعتماد به‌صورت تقریبی بر این فرض پیش‌بینی می‌شود که کاربران به شخص یا افرادی که درجه اعتماد بالاتری در شبکه دارند نزدیک‌تر می‌شوند (Ziegler and Lausen, ۲۰۰۴, Zenebe and Norcio, ۲۰۰۹).

علاوه بر اعتماد، موارد انبوه دیگری از رابطه‌های اجتماعی نیز وجود دارد که می‌تواند برای پیشنهادگرها استفاده شود. به‌طور مثال برگزیده‌های اجتماعی<sup>۳۶</sup>، بافت فیزیکی، برچسب‌های اجتماعی، هم‌نویسندگی<sup>۳۷</sup> و موارد بسیار دیگر که اخیراً برای استفاده در سامانه‌های پیشنهادگر به‌جای استفاده صرف از اعتماد در شبکه و یا معیارهای شباهت استفاده شده است. به‌طور مثال (Zenebe and Norcio, ۲۰۰۹) از یک سامانه برخط پیشنهادگر نام می‌برند که بر اساس شبکه هم‌نقل<sup>۳۸</sup> کردن از موارد پسند شده برخط توسعه داده است که موارد هم‌نقل شده از موارد موردپسند شده کاربران به‌عنوان وزن در نظر گرفته شده است.

---

<sup>۳۵</sup> Social network-based recommendation techniques

<sup>۳۶</sup> Social bookmark

<sup>۳۷</sup> co-authorship

<sup>۳۸</sup> Co-cited

پیشنهادگرهای مبتنی بر بافت آگاهی<sup>۳۹</sup> یکی از پرارجاع‌ترین تعاریف از بافت، تعریف دی و همکاران است که بافت را هرگونه اطلاعاتی که می‌تواند برای مشخص کردن وضعیت یک نهاد استفاده کرد گویند. یک نهاد می‌تواند یک شخص، یک مکان یا یک شی که می‌تواند مرتبط با تعامل بین کاربر و برنامه باشد در نظر گرفته شود که خود می‌تواند شامل کاربر یا یک برنامه باشد. اطلاعات بافت از قبیل زمان، اطلاعات جغرافیایی و یا همراهان فرد (به‌طور مثال دوستان، خانواده و یا همکاران) اخیراً در سامانه‌های پیشنهادگر موجود به کار گرفته شده‌اند. به‌طور مثال، اطلاعاتی که از دستگاه‌های موبایل بدست می‌آید مورد استفاده قرار می‌گیرد. اطلاعات بافت، اطلاعات اضافه‌ای را برای تولید پیشنهادها مناسب‌تر تولید می‌کند به‌خصوص برای برنامه‌هایی که اطلاعات کاربر و اقلام و موارد برای ارائه پیشنهاد کافی به نظر نمی‌رسد، به‌طور مثال برنامه‌ریزی سفر یا شخصی سازی اطلاعات یک وب سایت برای یک کاربر از این موارد هستند. همچنین ضروری است تا اطلاعات بافت در طی فرآیند ارائه پیشنهاد در نظر گرفته شود تا بتوان به یک کاربر در شرایط خاص پیشنهادها مناسب‌تری ارائه کرد، به‌طور مثال با در نظر گرفتن شرایط آب و هوایی یک پیشنهادگر سفر می‌تواند پیشنهادها متفاوت‌تری در تابستان نسبت به زمستان به کاربر ارائه دهد. برای وارد کردن اطلاعات بافت به یک پیشنهادگر، مقاله (Adomavicius and Tuzhilin, ۲۰۱۱) یک فرآیند سه مرحله برای محاسبه و ارزش‌گذاری اطلاعات ارائه داده است. مراحل به ترتیب بافت قبل از پالایش، بافت بعد پالایش و مدل بافت است. با پردازش این سه مرحله، سامانه می‌تواند بررسی کند که آیا اطلاعات بافت مناسب و سودمند برای پیشنهاد هستند. (Zenebe and Norcio, ۲۰۰۹)

روش پیشنهادگرهای گروهی<sup>۴۰</sup> پیشنهادگرهای گروهی، برای ارائه پیشنهادها به دسته‌ای از کاربران است که نمی‌توانند که دورهم مذاکره کنند و یا تمایلات آن‌ها به‌طور شفاف مشخص نیست. این پیشنهادگرها به گروه‌بندی الکترونیکی<sup>۴۱</sup> هم مشهور هستند و برای حوزه مختلفی مانند فیلم، گردشگری، موسیقی و صفحات وب مورد استفاده قرار می‌گیرند. راهبردهای متفاوتی با الهام از تئوری انتخاب‌های اجتماعی و فرآیند تصمیم‌گیری، برای رسیدن به یک جمع‌بندی در یک گروه استفاده می‌شود. جدا از رویکرد

---

<sup>۳۹</sup> Context awareness-based recommendation techniques

<sup>۴۰</sup> Group recommendation technique

<sup>۴۱</sup> E-group

جمع‌بندی، رویکردهای ارتباطی متقارن<sup>۴۲</sup> و غیرمتقارن<sup>۴۳</sup> در این سامانه‌ها استفاده می‌شود. (Lu et al., ۲۰۱۵)

رویکرد مبتنی بر کلیشه‌ها<sup>۴۴</sup> یکی از جدیدترین روش‌ها مدل‌سازی کاربر و دسته پیشنهادگرها هستند که توسط ریچی در سامانه پیشنهادگر گروندی<sup>۴۵</sup> برای پیشنهاد رمان به کاربران طراحی و استفاده شد. ریچی کلیشه‌ها را از روانشناسان که افراد را بر اساس ویژگی‌های شخصیت‌شان ارزیابی می‌کنند، الهام گرفت. ریچی کلیشه‌ها را مجموعه‌ای از ویژگی‌ها نامید. به‌طور مثال، گروندی فرض می‌کند که کاربران مذکر، طاقت بیشتری در برابر خشونت دارند و هیجان و تعلیق، خط داستانی سریع را ترجیح می‌دهند و نسبت به رمان‌های عاشقانه نظر منفی دارند؛ بنابراین گروندی کتاب‌هایی را به کاربر پیشنهاد می‌کند که با ویژگی‌ها و جنبه‌ها شخصیتی کاربر تطابق دارد.

یکی از مشکلات عمده در رابطه با کلیشه‌ها، طبقه‌بندی سخت آن‌ها در رابطه با کاربران است. به‌طور مثال تمام مردان نسبت به رمان‌های عاشقانه نظر منفی ندارند. مشکل دیگر اختصاص اقلام و مواردی پیشنهادی به هر کدام از ویژگی‌های شخصیتی باید به‌صورت دستی انجام گیرد که برای اقلام متنوع و زیاد مشکل به نظر می‌رسد؛ اما در مقابل انتقاد از این روش، یکی از مزایای این روش پردازش کم کامپیوتری برای ارائه پیشنهاد و همچنین نتایج مناسب این روش در ارائه پیشنهاد و رضایت کاربران است. به‌طور مثال بسل و همکاران از قول و بر و کاستیلو نقل می‌کند به‌طور معمول زمانی که زن‌ها در بخش جستجو یاهو نام واگنر را جستجو می‌کنند منظورشان ریچارد واگنر آهنگساز و موسیقیدان آلمانی است اما در مقابل زمانی مردان در یاهو واگنر را جستجو می‌کنند منظورشان اسپری نقاشی واگنر است. همچنین آژانس مسافرتی اوربیتز دریافت کرده است که کاربران با کامپیوترهای مک شرکت اپل، احتمال رزرو هتل‌های چهار و پنج ستاره ۴۰٪ بیشتر از سایر پلتفرم‌های دیگر است و زمانی که یک هتل را انتخاب می‌کنند به‌طور معمول اتاق‌های گران‌تر را انتخاب می‌کنند؛ بنابراین یک کلیشه از کاربران مک و سایر پلتفرم‌ها ساخته‌اند و زمانی که کاربری با کامپیوتر مک به وب‌سایت آن‌ها مراجعه می‌کند موارد گران‌تری به آن‌ها پیشنهاد می‌شود. در زمینه‌های پیشنهادگرهای مقالات و کتب علمی، بیل و همکارانشان نیز از کلیشه‌ها برای ارائه پیشنهادها استفاده کرده‌اند. کلیشه آن‌ها بر اساس این است که کاربر آن‌ها دانشجویان یا محققین هستند و بنابراین مواردی را پیشنهاد می‌کنند که خوشایند کاربر باشد (Beel et al., ۲۰۱۶, Weber and Castillo, ۲۰۱۰).

---

<sup>۴۲</sup> Synchronous

<sup>۴۳</sup> Asynchronous

<sup>۴۴</sup> Stereotypes

<sup>۴۵</sup> Grundy

- آنتولوژی<sup>۴۶</sup> یک آنتولوژی تشریح کننده یک دامنه مشترک و به طور روشن مفهومی شده است. عناصر اصلی آن شامل دسته‌ها (بیانگر مفاهیم که معمولاً در یک ساختار سلسله مراتبی قرار می‌گیرد)، ضوابط طبقه‌بندی و غیر طبقه‌بندی شده، قواعد و در آخر نمونه‌ها است.
- آنتولوژی دارای دسته‌های مختلفی بر اساس کاربرد است که شامل موارد زیر است:
- آنتولوژی برنامه‌ها: که توسط برنامه‌ها مورد استفاده قرار می‌گیرد به طور مثال آنتولوژی فرآیند تولید محصول آنتولوژی حفره‌های امنیتی و غیره.
  - آنتولوژی دامنه: این آنتولوژی‌ها برای نمایش دانش در دامنه خاص است.
  - آنتولوژی فن‌های پایه‌ای: تشریح کننده ویژگی‌های عمومی یک ابزار و یا وسیله است به طور مثال شامل اجزا فرآیندها و عملگرها است.
  - آنتولوژی‌های عمومی و کلی: ارائه جزئیات از مفاهیم کلی مانند زمان و فضا و غیره می‌دهد.
  - آنتولوژی دارای عناصر زیر است که می‌تواند برای نمایش دانش در هر زمینه خواص مفید باشد.
  - مفاهیم: ایده اصلی است. مفاهیم می‌توانند نوعی از فن‌ها، روش‌ها، راهبردها، فرآیندهای منطقی و غیره باشد.
  - ارتباطات: نمایانگر تعامل و پیوندهای میان مفاهیم در یک زمینه است آن‌ها به طور معمول شکل‌دهنده حالت سلسله مراتبی یک زمینه هستند.
  - عملگر: نوع به هم پیوسته ارتباطات است جایی که یک عنصر با محاسبه یک عملگر که چندین عنصر از یک آنتولوژی را در نظر می‌گیرد شناخته می‌شود.
  - نمونه‌ها: برای نمایش یک شیء به خصوص از یک مفهوم استفاده می‌شود.
  - قواعد: نوع ارتباطی را که عناصر یک آنتولوژی باید داشته باشند را نمایش می‌دهد. به طور مثال اگر الف و ب زیرمجموعه و تشکیل‌دهنده ج هستند بنابراین الف زیرمجموعه ب نیست (Moreno et al., ۲۰۱۳, Ciurana Simó, ۲۰۱۲).

## ۲,۳,۵ روش‌های ارزیابی سامانه‌های پیشنهادگر و بسندگی آن‌ها

برای ارزیابی میزان مؤثر بودن یک سامانه پیشنهادگر روش‌ها و فن‌های متنوعی وجود دارد که هر کدام دارای مزایای و معایب خود هستند. در مقاله مروری (Beel et al., ۲۰۱۶)، بیل و همکاران بیان می‌دارند از ۹۶-مقاله بررسی شده توسط آن‌ها در زمینه‌ی پیشنهادگرها، نزدیک به ۲۲٪ آن‌ها توسط نویسندگان، در مقالاتشان ارزیابی نشده است. همچنین بیان می‌دارند که برخی دیگری از این مقالات دارای روش ارزیابی

---

<sup>۴۶</sup> Ontology

مناسب نمی‌باشند و معتقد هستند که روش آن‌ها غیر شفاف است. در این قسمت سعی بر این است که روش‌های ارزیابی سامانه‌های پیشنهادگر، بررسی و تحلیل شود.

**مطالعه کاربران**<sup>۴۷</sup> مطالعه کاربران، به‌طور معمول میزان رضایت کاربران را با دریافت رتبه‌دهی صریح از آن‌ها، اندازه‌گیری می‌کند. کاربران پیشنهادهای متفاوتی که از پیشنهادگرهای مختلف به آن‌ها می‌شود را امتیازدهی می‌کنند بنابراین پیشنهادگری که بیشترین امتیاز را دریافت کند، عملکرد بهتری داشته و میزان رضایت بیشتری از سوی کاربر داشته است. به‌طور معمول در این رویکرد، سؤالات برای دریافت بازخورد و امتیاز از کاربر از جنبه‌های عمومی پیشنهادگر است اما ممکن است به جنبه‌های خاصی از پیشنهادگر در سؤال نیز اشاره شود و نظر کاربر را جویا شوند. به‌طور مثال، پرسیده شود چقدر پیشنهاددهی معتبر و یا نو هستند و یا چقدر برای افراد غیرمتخصص مناسب است. ممکن است در این رویکرد بازخوردهای کیفی نیز دریافت شود و اما این نوع بازخوردها به‌ندرت در پیشنهادگرها استفاده می‌شود.

به‌طور کلی مطالعات کاربری دو نوع مطالعه آزمایشگاهی که کاربر اطلاع از این دارد که پیشنهادگر در حال آزمایش است و از او خواسته می‌شود به پیشنهادگر امتیاز بدهد و نوع دیگر دنیای واقعی است که در آن کاربر، بی‌اطلاع از این است که پیشنهادگر موردنظر تحت آزمایش و مطالعه است. بیل و همکاران از قول (Cremonesi et al., ۲۰۱۲) بیان می‌دارند که نوع سؤالات پرسیده شده، در دقت و نتیجه مطالعه اثربخشی پیشنهادگر بسیار تأثیرگذار است. همچنین پیدا کردن افراد داوطلب برای مطالعه اثربخشی پیشنهادگر، یکی از دشواری‌های این روش به شمار می‌رود، اما برای یک نتیجه مناسب حداقل باید ده‌ها فرد با پیشنهادگر تعامل داشته و بازخورد آن‌ها بررسی شود (Beel et al., ۲۰۱۶).

**ارزیابی برخط**<sup>۴۸</sup> ارزیابی برخط ابتدا توسط حوزه‌های تبلیغات آنلاین و تجارت الکترونیک استفاده شد. آن‌ها نرخ پذیرش پیشنهادها در دنیای واقعی پیشنهادها را اندازه‌گیری می‌کردند. نرخ اندازه‌گیری عموماً از طریق کلیک<sup>۴۹</sup> یا به‌طور خلاصه CTR اندازه‌گیری می‌شود. به‌طور مثال نرخ کلیک‌های یک پیشنهادگر برای نمایش پیشنهادها از این موارد است. به‌طور مثال اگر یک پیشنهادگر، ۱۰۰۰۰ پیشنهاد نمایش دهد و ۱۲۰ کلیک صورت گیرد، گفته می‌شود CTR آن برابر با ۱,۲٪ است. معیارهای اندازه‌گیری دیگری مانند نرخ بازگیری یا خرید اقلام هم وجود دارد. نرخ پذیرش به‌طور معمول به‌عنوان یک معیار غیرصریح برای سنجش سطح رضایت کاربر مطرح می‌شود. اگرچه این معیارها خیلی قابل‌اعتماد نیستند به‌طور مثال

---

<sup>۴۷</sup> User studies

<sup>۴۸</sup> Online evaluation

<sup>۴۹</sup> Click through rates(CTR)

ممکن است کاربری کتابی را خریداری کن و بعد از خواندن آن، نمره منفی دهد. از این جهت برای استفاده از این روش‌ها باید احتیاط به خرج داد و همچنین این روش‌ها نسبت به روش‌های غیر برخط زمان‌بر و دشوارتر و پرهزینه‌تر هستند (Beel et al., ۲۰۱۶).

**ارزیابی برون خطی**<sup>۵۰</sup> ارزیابی برون خطی میزان دقت یک پیشنهادگر را بر اساس صحت پیشنهادها ارائه شده توسط آن اندازه‌گیری می‌کند. برای اندازه‌گیری دقت، بررسی می‌شود که در چند پیشنهاد برتری که ارائه شده چه تعداد از آن‌ها به درستی پیشنهاد شده‌اند. معیارهای اندازه‌گیری دیگر شامل نسبت بازیافت<sup>۵۱</sup>، اندازه‌گیری-ف<sup>۵۲</sup>، میانگین امتیاز متقابل<sup>۵۳</sup>، میانگین مطلق خطا<sup>۵۴</sup> و جذر میانگین مربع خطا<sup>۵۵</sup> است. روش ارزیابی برون خط در اصل نویدبخش شناسایی تعدادی از رویکردهای پیشنهادگرها شد. این روش به‌طور معمول به همراه مطالعه کاربران یا مطالعات برخط برای بررسی اثربخش بودن پیشنهادگرها استفاده می‌شود. چندین پژوهش نشان داده است که نتایج ارزیابی برون خط با نتایج رویکرد مطالعات کاربران و یا ارزیابی برخط، لزوماً منطبق نیست. یعنی می‌توان نتیجه گرفت که موفقیت یک پیشنهادگر با ارزیابی برون خطی تضمین‌کننده موفقیت آن در عمل و دنیای واقعی نیست. از جمله دلایل انتقادات به ارزیابی برون خطی، این است که این روش ویژگی‌های انسانی در رضایت آن از یک پیشنهاد را در نظر نگرفته و صرفاً به بررسی و شمارش پیشنهادها درست می‌پردازد اما باین وجود معیار رایج و غالب ارزیابی پیشنهادگرها است. بیل و همکاران بیان می‌دارند که اختلافات در رابطه با ارزیابی برون خطی و مطالعات کاربری علاوه بر دنیای واقعی در حوزه پژوهش‌های پیرامون زمینه پیشنهادگرها نیز است. به‌طور مثال دو مقاله (Ekstrand et al., ۲۰۱۰, Liang et al., ۲۰۱۱) نتایج ارزیابی برون خطی آن‌ها شبیه به ارزیابی مطالعه کاربران بود. اگرچه جامعه مورد آزمایش این دو مقاله پنج و ۱۹ نفر بودند که از لحاظ آماری نتایج آن معنی‌دار نیست. و در ادامه از مقالات دیگر نام می‌برد که نتایج ارزیابی برون خط با نتایج مطالعه کاربر در آن‌ها (جامعه مورد آزمایش دو مورد از آن‌ها بیش از ۱۰۰ نفر بودند) متضاد بودند. و نتیجه‌گیری می‌کند که ارزیابی برون خطی به‌طور قابل اطمینان نمی‌تواند اثربخشی پیشنهادگر را در دنیای واقعی پیش‌بینی کند (Beel et al., ۲۰۱۶).

البته نوع دیگری از ارزیابی موسوم به مبتنی بر سناریو است که یک کاربر فرضی را در نظر گرفته و تمام مراحل الگوریتم پیشنهادگر را با آن کاربر طی می‌کنند. به‌طور معمول زمانی از این نوع ارزیابی استفاده

---

<sup>۵۰</sup> Offline evaluation

<sup>۵۱</sup> Recall

<sup>۵۲</sup> F-measure

<sup>۵۳</sup> Mean reciprocal rank(MRR)

<sup>۵۴</sup> Mean Absolut error

<sup>۵۵</sup> Root mean square error



می شود که پژوهشگر ادعایی در رابطه با رفع مشکلات جاری پیشنهادگرها با الگوریتم طراحی شده خود دارد.

#### ۲,۳,۶ مشکلات و چالش های سامانه های پیشنهادگر

هرکدام از رویکردها برای طراحی سامانه های پیشنهادگر دارای مزایا و معایبی هستند که به طور خلاصه در مطالب گذاشته هم به آنها پرداختیم. به طور کلی طراحی پیشنهادگرها دارای مشکلاتی است که در قسمت زیر به آنها پرداخته ایم.

**مشکل صعود:** این مشکل خود دربرگیرنده دو مشکل دیگر است، مشکل ورود کاربران تازه، زمانی که یک کاربر وارد سیستم می شود باید رتبه دهی بسیاری کند تا سامانه از او شناخت پیدا کرده و به او پیشنهادات مناسب ارائه دهد. برای ارائه پیشنهادات مناسب، نیاز است میزان امتیازدهی و پاسخ کاربر به سؤالات زیاد باشد. اگر میزان این امتیازدهی های کاربر در ابتدا ورود به سامانه به اندازه کافی نباشد، پیشنهادات ارائه شده هم مناسب نخواهد بود. این مشکل در رویکردهای پالایشگرهای مبتنی بر محتوا و پالایشگرهای مشارکتی اتفاق می افتد. این مشکل در نقطه مقابل می تواند برای اقلام و موارد جدید که برای پیشنهاد وارد سامانه می شوند بیفتد. به این دلیل سامانه احتمالاً اقلام جدید را از این جهت که نرخ امتیاز گیری کمتری نسبت به سایر اقلام دارد، به کاربران پیشنهاد نمی کند. این مشکل در سامانه های پالایشگر مشارکتی مشاهده می شود. نام دیگر این مشکل که قبلاً نیز بدان پرداختیم، مشکل شروع سرد است.

**مشکل تنک بودن داده ها:** امتیازات که کاربر حاضر به اقلام می دهد ممکن است با امتیازاتی که دیگر کاربران می دهند تطابق نداشته باشد. این مشکل نیز در سامانه های پالایشگر مبتنی بر محتوا و پالایشگرهای مشارکتی مشاهده می شود.

**مشکل در دریافت اطلاعات دموگرافیک:** به طور کلی دریافت اطلاعات دموگرافیک برای کاربر بسیار حساس برانگیز بوده و ممکن است در اعتماد و تجربه کاربر تأثیرگذار باشد. یکی از راه حل ها برای پرداخت به این مشکل، تعیین حداقل اطلاعات دموگرافیک است که از کاربر درخواست می شود.

**اثر پورتفلو<sup>۵۶</sup>:** پیشنهادگرها نباید پیشنهادهایی به کاربر بدهند که این پیشنهادات تفاوتی چندانی باهم نداشته و شباهت زیادی به هم داشته باشند. به طور مثال در یک پیشنهادگر کتاب، پیشنهادگر نباید چاپ جدید یک کتاب را که در گذشته کاربر خوانده است را پیشنهاد کند.

---

<sup>۵۶</sup> Portfolio

پیشنهادها یا خیلی کم یا خیلی زیاد: انتخاب تعداد مناسب پیشنهادها برای کاربران، مسئله جدی است. پیشنهادها کم ممکن است منجر به انتخاب و پیدا کردن پیشنهاد مناسب با کاربر نشود و از طرف دیگر پیشنهادها زیاد می‌تواند مجدداً منجر به افزونگی اطلاعات و پیچیدگی مسئله برای کاربر شود. تکرار گرایی و شخصی‌سازی بیش‌ازحد<sup>۵۷</sup>: برخی از پیشنهادگرها فقط اقلامی را به کاربر پیشنهاد می‌کنند که مشابهشان توسط کاربر در گذشته پسندیده شده‌اند؛ اما باید نوگرایی در پیشنهادات وجود داشته باشد، ممکن است کاربر به پیشنهادات نو و متفاوت از پیشنهادات و موارد پسندیده شده علاقه نشان دهد (Ricci et al., ۲۰۱۵).

گوسفند خاکستری<sup>۵۸</sup>: منظور کاربر یا اقلامی است که شباهتی با کاربر و یا اقلام دیگر ندارد یا به عبارتی اشتراکات آن‌ها حداقل است. این مشکل در پیشنهادگرها مبتنی بر مشارکت اتفاق می‌افتد.

## ۲.۴ پیشنهادگرها و گردشگری

بنا بر گزارش کمیسیون گردشگری اتحادیه اروپا در سال ۲۰۰۹، تقریباً دوسوم از مردم دنیا از اینترنت برای پیدا کردن مقاصد گردشگری خود استفاده می‌کردند که امروزه با توجه به ضریب نفوذ اینترنت و راحتی دسترسی به آن، این میزان افزایش پیدا کرده است. انتخاب مقاصد گردشگری علاوه بر این که فرآیندی احساسی است، دارای جنبه‌ها و ویژگی‌های متنوعی نیز است که این به پیچیدگی مسئله دامن می‌زند. به دلیل وجود داده‌های فراوان در دنیای وب، افزونگی اطلاعات وجود دارد که این خود موجب مشکلی جدی در بدست آوردن اطلاعات مناسب گردشگری با توجه به نیازها و محدودیت‌های کاربر می‌شود. از طرف دیگر، کاربران عادی دارای دانش زمینه نیستند و نمی‌توانند نیازها و خواسته‌های خود را به درستی جستجو کنند بنابراین در این میان برای رفع این مشکلات طرف‌های ارائه‌دهنده خدمات گردشگری، به سراغ سامانه‌های پیشنهادگر برای افزایش تجربه کاربری مشتریان و خدمات گیرندگان می‌روند (Neidhardt et al., ۲۰۱۵).

بنا به استدلال بروک و رمضانی، پیشنهادگرهای مناسب در حوزه گردشگری، مبتنی بر محتوا و مبتنی بر دانش هستند. این استدلال از آنجا ناشی می‌شود که گردشگری یک حوزه پرمخاطره (به دلیل گران بودن تفریحات و سفر نسبت به موارد دیگر)، تغییرات کم (مقدار و ارزش موارد گردشگری تغییر نمی‌کند)، تنوع کم، ترجیحات غیر پایدار کاربران (ترجیحات کاربران در گذشته ممکن است با ترجیحات امروز آن‌ها

---

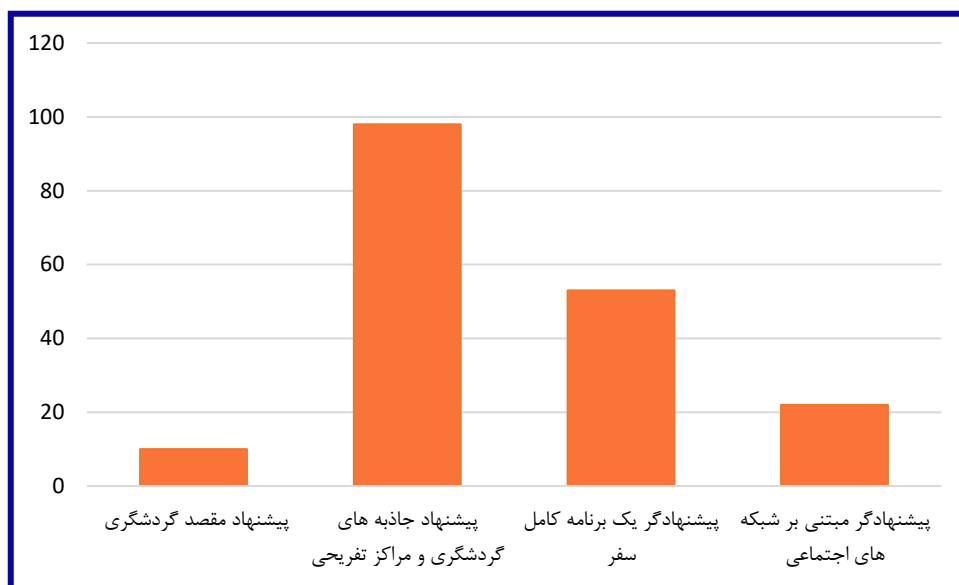
<sup>۵۷</sup> Over specialization

<sup>۵۸</sup> Grey sheep

یکسان نباشد) و تعاملات صریح کاربران (کاربران نیاز به فرموله کردن نظراتشان و یا انجام جستجو به منظور افزودن داده‌های شخصی‌شان دارند)، است (Burke and Ramezani, ۲۰۱۱).

در نتیجه، پیشنهادگرهای حوزه گردشگری در قیاس با پیشنهادگرهای کتاب، موسیقی و غیره بسیار پیچیده‌تر و نامحسوس‌تر است؛ بنابراین برای طراحی یک پیشنهادگر، باید جنبه‌های مختلف گردشگری مانند جاذبه‌های گردشگری، مسیر و حمل و نقل مناسب و به‌طور کلی یک مجموعه مناسب برای گردشگر ارائه شود. برای طراحی چنین پیشنهادگری نیاز به دانش و محتوا وجود دارد. به‌طور مثال زمان پرواز و زمان رزرو هتل و میزان روزهای اسکان شخص در هتل، به هم وابسته است و باید در پیشنهادگر لحاظ شود.

گارسا و همکاران در پیشنهادگرهای گردشگری دو گونه کلی از آن‌ها را شناسایی کرده‌اند. در نوع اول، پیشنهادگرها یک مقصد خاص و مجموعه‌ای از تفریحاتی که می‌شود انجام داد را پیشنهاد می‌کند. به‌طور مثال ممکن است، باغ فردوس در استان تهران و به همراه گشت و گذار در موزه سینما و کافه‌های آن را پیشنهاد کند. نوع دیگر در مقصدی خاص، با توجه به نیازها و محدودیت‌های کاربر پیشنهادات ارائه می‌کند. به‌طور مثال در شهر تهران، ممکن است به کاربر بازدید از موزه و تماشاخانه یک رویداد ورزشی را پیشنهاد کند (Garcia et al., ۲۰۱۱). البته عملکرد پیشنهادگرها محدود به این دو مورد نمی‌شود که البته می‌شود آن‌ها را زیرمجموعه این دو و یا ترکیبی از این دو دانست. به‌طور مثال در نوع دیگر پیشنهادگرها برنامه‌ریزی یک سفر کامل همراه با مسیرهای گردشگری را انجام می‌دهند و یا در نوع دیگر که مبتنی بر شبکه‌های اجتماعی هستند، گردشگران تصاویر و تجربیات خود را با سایرین به اشتراک می‌گذارند و مقصد یا برنامه گردشگری خاصی را به هم پیشنهاد می‌کنند. شکل زیر مقایسه‌ای تصویرگون از میزان پژوهش‌های انجام‌شده پیرامون هر کدام از انواع این پیشنهادگرهای گردشگری است.



شکل ۲-۲ مقایسه فراوانی مقالات نوع پیشنهادگرهای گردشگری بر اساس عملکرد (Borràs et al., ۲۰۱۴)

#### ۲،۴،۱ نقش ها و شخصیت ها در پیشنهادگرهای گردشگری

از سال های ۱۹۷۰ به این سو پژوهشگران سعی در ایجاد پیوند بین رفتارهای گردشگری و جنبه های روان شناختی کاربران داشتند. یکی از تأثیرگذارترین پژوهش ها در این حوزه، پژوهش ها یاناکیز و گیبسون است. در پژوهش های شان، ۱۵ نقش پیش تعریف شده، با ۳۰ سؤال در پرسشنامه در ارتباط با رفتارهای گردشگری که هر دو سؤال برای یک نقش است، تعدادی از افراد را مورد ارزیابی قرار دادند. مجموعه ای از نیازمندی ها شامل نیاز برای خانه و خانواده، نیاز به کنترل شرایط، نیاز به امنیت و همراه، را لیست کردند. همچنین ویژگی های دموگرافیک شامل سن و جنسیت در این پژوهش پرسیده شد (Neidhardt et al., ۲۰۱۵). مجدداً یاناکیز و گیبسون، بر مبنای مطالعات قبلی خود در سال ۲۰۰۲ نتایج پژوهش های جدید خود را منتشر کردند. هدف پژوهش های آنها بررسی ارتباط نقش های گردشگری و نیازهای روان شناختی افراد (مرد و زن) در سراسر طول دره زندگی آنها بود. شواهد آماری نشان دهنده این است که الگوی رفتارهای گردشگری به نیازهای روان شناختی افراد وابسته بوده و دارای ارتباط هستند و این مهم در طول زندگی آنها تغییر می کند. آنها همچنین به ترتیب دو نقش گردشگری، گردشگر مستقل و گردشگر فراری را به دو نقش گردشگر مستقل اول و گردشگر مستقل دوم، گردشگر فراری اول و گردشگر فراری دوم افزایش دادند و در مجموع ۱۵ نقش گردشگری اولیه به ۱۷ نقش گردشگری تبدیل شد که در جدول ۲-۳ آمده است (Gibson and Yiannakis, ۲۰۰۲). در پژوهش (Berger et al., ۲۰۰۷)، در یک پرسشنامه ۱۷-نقش گردشگری به طور غیر صریح و با بیان جملات توصیفی از این نقش ها، از کاربر سؤال شد که کدامیک از این نقش ها رفتار او را در گذشته و حال توصیف می کند. سپس در بخش دوم پرسشنامه از

کاربر خواسته می‌شود تا از ۶۰ عکس ارائه شده که بیانگر ده وضعیت متفاوت گردشگری بود، خواسته شد که عکس‌هایی را که بیانگر رفتار گردشگری او در حال حاضر و گذشته است را برگزیند. همچنین اطلاعات شخصی مانند اطلاعات ویژگی‌های دموگرافیک شخص از او سؤال می‌شد. نتایج بیانگر این بود که ویژگی‌ها و نقش‌ها شخصیتی به کمک تصاویر قابل دریافت و شناسایی هستند. همچنین این پژوهش بیان می‌کرد که برخی از نقش‌ها شبیه یکدیگر بوده و تفکیک آن‌ها از یکدیگر دشوار است.

همچنین مقاله (Gretzel et al., ۲۰۰۴) یک مطالعه جهت نمایش ارتباط شخصیت گردشگری و رفتارهای گردشگری انجام داده است. در این مطالعه که در شمال آمریکا انجام شده است، از کاربران خواسته شده تا به اهمیت انگیزه‌های سفری مشخص شده برای آن‌ها (به‌طور مثال ارتباطات اجتماعی، فعالیت‌های فیزیکی، پیدا کردن آرامش) امتیاز دهند. البته سؤالات دیگری در رابطه سبک گردشگری و ارزش گردشگری از آن‌ها سؤال شد. برای فهم رفتار گردشگری واقعی کاربر، از شرکت‌کنندگان خواسته شد، که دقیقاً مقصد و فعالیت‌های گردشگری اخیر آن‌ها در ناحیه در نظر گرفته شده را مشخص کنند. بعلاوه دوازده نوع شخصیت گردشگری با کمک‌گیری از وب‌سایت‌های گردشگری و بر مبنای فعالیت و جاذبه‌های گردشگری تعریف کرده است. نتایج پژوهش‌ها حاکی از این است که شخصیت‌ها از روی سبک گردشگری، انگیزه گردشگری و ارزش سفر برای شخص قابل تشخیص و دریافت است. همچنین دریافت که اگر، محدودیت انتخاب یک مورد برداشته شود، کاربران مایل هستند تا تمام مواردی را که توصیف‌گر شخصیت آن‌ها است را انتخاب کنند. همچنین یک رابطه نزدیک بین شخصیت گردشگری و فعالیت‌ها موردپسند کاربران وجود دارد. اگرچه در مطالعات هیچ‌گونه ارتباط محسوسی بین انتخاب مقصد گردشگری و شخصیت گردشگری پیدا نشد. نتیجه کلی این بود که انتخاب مقصد گردشگری هیچ‌وقت به کمک پیشنهادگرهای که از دسته‌های شخصیتی استفاده می‌کنند بدست نمی‌آید. یکی از وب‌سایت‌های گردشگری که از دسته‌های شخصیت‌های گردشگری استفاده کرده است ریزولتون تی یو آی<sup>۵۹</sup> است که دارای هفت دسته‌بندی مختلف است که برای هر کدام تعدادی اختصاص داده است. بعد از پاسخ به سه سؤال و انتخاب بیش از هشت اولویت مانند ورزش، سرگرمی، اینترنت بی‌سیم و غیره، یک یا چند نوع از این دسته‌بندی‌ها به کاربر پیشنهاد می‌شود. همچنین انتشارات کتاب‌های راهنمای سفر لائلی پلنت<sup>۶۰</sup>، یک کمپین برای سازمان خدمات گردشگری آمریکا توسعه داده است. نوع شخصیت گردشگری شخص با پرسش چهار سؤال و سه پاسخ محتمل برای هر کدام از آن‌ها مشخص می‌شود. پنج نوع شخصیت گردشگر متفاوت با نام‌های، گشت‌وگذار جهانی، عاشق خرید، جهانگرد سنتی، هیجان‌طلب و فضای نوظلب است. البته در هر دو مورد تجاری گفته شود، هیچ‌گونه اشاره‌ای به نحوه پیدا کردن و ارائه شخصیت‌های گردشگری نداده‌اند.

<sup>۵۹</sup> Reisewelten TUI

<sup>۶۰</sup> Lonely plant

به طور کلی می توان نتیجه گیری کرد، یک پیشنهادگر مبتنی بر نمایه و ویژگی های کاربر بادید دو جنبه زیر را رعایت کند:

توانایی دریافت اطلاعات و ترجیحات کاربر، به روشی غیر از بیان جملات زیاد و طولانی باشد.  
ارائه یک روش محاسباتی مناسب که بر مبنای تشریح ترجیحات کاربر و تعلق آن به رویکردهای مبتنی بر دانش و مبتنی بر محتوا باشد (Neidhardt et al., ۲۰۱۵).

جدول ۳-۲ نقش های گردشگری هفده گانه (Gibson and Yiannakis, ۲۰۰۲)

نقش گردشگری	توصیف نقش گردشگری
۱	علاقه مند به آرامش در گرما و شن های ساحل
۲	علاقه به مهمانی و رفتن به کافه ها و ملاقات افراد جدید
۳	علاقه مند به ملاقات افراد محلی، امتحان کردن غذاها و صحبت به زبان آن ها
۴	علاقه به مطالعه تاریخ و بازدید از کاخ ها و ویرانه های تاریخی
۵	علاقه به سفرهای برنامه ریزی شده و خرید و عکس انداختن
۶	علاقه به مخاطره و شرکت در فعالیت های که برای آن ها پرهیجان باشد
۷	علاقه به سفر ماجراجویانه، کشف مکان ها نو و چالش های موجود آن
۸	تعطیلات در نقاط ممتاز، رفتن به کلوب های شبانه خاص و هم نشینی با افراد مشهور
۹	جستجو به دنبال شناخت خود (روح) یا افزایش دانش برای درک معنی زندگی
۱۰	بازدید از مکان ها و جاذبه های معمول و پرهیز سفرها و مجموعه های برنامه ریزی شده
۱۱	انتخاب مقصد و رزرو هتل را خود به تنهایی و اغلب با شنیدن توصیه دیگران
۱۲	سفر با بلیت سطح اول، بهترین هتل ها، رفتن به نمایش ها و لذت بردن از شام عالی
۱۳	یک نوع هپی، رانندگی از محلی به محلی دیگر
۱۴	از خلاص شدن از استرس و محیط پرفشار خانه لذت می برد
۱۵	خلاص شدن از تمام چیزها با فرار به یک منطقه آرام، خالی از سکنه
۱۶	در طول سفر به فعالیت و انجام ورزش های مورد علاقه تأکید دارد
۱۷	در سفرهای مطالعاتی شرکت می کند و به هم اندیشی ها برای کسب مهارت و دانش می رود

## ۲،۴،۲ مروری بر پژوهش های پیشین پیشنهادگرهای گردشگری

حضور و نفوذ گسترده اینترنت با رواج دستگاه های قابل حمل مانند تلفن های همراه و تبلت ها بیش از گذشته شده است. علاوه بر اینترنت وجود تراشه های موقعیت سنج مکان کاربر نیز منجر به تنوع و کارآمدی پیشنهادگرها شده است. بورک و همکاران (Burke et al., ۱۹۹۶) یک پیشنهادگر رستوران به نام انتری<sup>۶۱</sup> با ترکیب رویکرد مبتنی بر محتوا و مبتنی بر دانش، طراحی کرده اند. دانش از سمت کاربر جمع آوری می شود و انتری آن را بازایی می کند تا انتخاب های مشابه بر اساس معیارهایی مانند قیمت و

<sup>۶۱</sup> entree

طعم پی‌شنهاد کند. مقاله (Tung and Soo, ۲۰۰۴) یک سامانه پی‌شنهادگر برای پی‌شنهاد رستوران به گردشگران در تایپه طراحی کرده است. اطلاعات بافت مانند، موقعیت مکانی و ویژگی‌های بی‌سیم تلفن همراه فرد، به صورت پویا توسط یک مدیریت کننده اطلاعات بافت جمع‌آوری می‌شود. پیشنهادهای بر اساس جستجو کاربر و اطلاعات بافت به او نمایش داده می‌شود. پی‌شنهادگر دیگر برای رستوران، مقاله (Martinez et al., ۲۰۰۹) است که رویکرد مبتنی بر دانش و مبتنی بر محتوا را با یکدیگر ترکیب کرده است. زمانی که سامانه بتواند یک نمایه بر اساس امتیازدهی‌های کاربر بسازد، پیشنهادهای بر اساس رویکرد مبتنی بر محتوا ارائه می‌شود و زمانی که اطلاعات کافی از کاربر موجود نیست، بر اساس منطق مبتنی بر مورد پیشنهادات را ارائه می‌دهد. برای گردش کردن در شهر پورتو در کشور پرتغال، مقاله (Lucas et al., ۲۰۱۳) یک پی‌شنهادگر برای پیدا کردن برنامه‌های گردشگری سفارشی شده به کاربر ارائه و طراحی کرده است. برای جلوگیری از مشکلات پی‌شنهادگرها مانند تنک بودن داده‌ها، پراکندگی آن‌ها، شروع سرد یک پی‌شنهادگر ترکیبی ارائه شده است. این پی‌شنهادگر رویکردهای مبتنی بر محتوا، مبتنی بر مشارکت اشتراکی را با یک روش خوشه‌بندی، دسته‌بندی‌های انجمنی و منطق فازی برای افزایش بهره‌وری و عملکرد سامانه پی‌شنهادگر خود ترکیب کرده است. سیگتور<sup>۶۲</sup> (Sieg et al., ۲۰۰۷) برای پی‌شنهاد فعالیت‌های گردشگری سفارشی‌شده مختص کاربر در شهر تاراگونا طراحی شده است. برای ارائه پیشنهادات مناسب، این پی‌شنهادگر چندین رویکرد و فن را باهم ترکیب کرده است. اطلاعات مورد استفاده در این پی‌شنهادگر شامل اطلاعات دموگرافیک، جزئیاتی که چهارچوب سفر را مشخص می‌کند، جنبه‌های ژئوگرافی، اطلاعاتی که به صورت صریح توسط کاربر ارائه شده است و اطلاعات غیرصریح که با بازخوردهای کاربر دریافت شده است. این پی‌شنهادگر، رویکردهای مبتنی بر محتوا و پالایشگر مشارکتی را باهم ترکیب کرده و از فن‌های یادگیری ماشین مانند خوشه‌بندی خودکار، مدیریت آنتولوژی استفاده کرده است و یک روش جدید جهت اندازه‌گیری شباهت کاربران، تعریف کرده است. موزه هوشمند (Ruotsalo et al., ۲۰۱۳)، یک پی‌شنهادگر مبتنی بر تلفن همراه است که به کاربران بر اساس مکان آن‌ها نقاط و مکان‌های گردشگری پی‌شنهاد می‌کند. در این پی‌شنهادگر، آنتولوژی و پالایشگر اطلاعات برای ارائه پیشنهادات شخصی سازی شده، توسعه داده شده است. اطلاعات بافت جمع‌آوری شده چه از طرف کاربر و چه به طور غیرمستقیم از طریق حسگرهای تلفن همراه شخص، به لایه مفاهیم آنتولوژی وارد می‌شود. چهارچوب پالایشگر موجب ارائه پرسشگر مبتنی بر آنتولوژی گسترده می‌شود. متعادل سازی متغیرها و خوشه‌بندی نتایج در نهایت موجب بهبود چشمگیر در دقت پی‌شنهادگر می‌شود. پی‌شنهادگر دیگر در پژوهش (Yang and Hwang, ۲۰۱۳) مبتنی بر تلفن همراه توسط ارائه شده است. در این سامانه، روش پالایشگر

---

<sup>۶۲</sup> Sig-tur

مشارکتی و ارتباط نفر به نفر تلفن همراه ترکیب و استفاده شده است. برای استفاده از اطلاعات دیگر گردشگران با علاقه‌مندی‌های مشابه، سه روش مبادله اطلاعات برای کاربران جهت مبادله امتیاز که به جاذبه‌های که دیدن کرده‌اند، ارائه کرده است. مولسکینگ<sup>۶۳</sup> (Avesani et al., ۲۰۰۵)، یک وب‌سایت است که به جامعه کاربران خود کمک می‌کند تا فعالیت‌های اسکی‌بازی خود را برنامه‌ریزی کنند. این سامانه به کاران اجازه می‌دهد تا نظرات خود را در رابطه محل‌های ویژه اسکی‌بازی اشتراک بگذارند و همچنین درجه اعتماد به هر کاربر نیز سنجیده شود. افرادی که به اسکی می‌روند، می‌توانند اطلاعات وضعیت برف و امنیت جاده‌ها را داشته باشند. مقاله (Fesenmaier et al., ۲۰۰۳) یک پیشنهادگر مبتنی بر منطق موردی طراحی کرده است که یک برنامه کامل گردشگری پیشنهاد می‌کند. کاربران به شیوه‌های مختلف می‌توانند از سامانه استفاده کنند. کاربر می‌تواند یک برنامه کامل سفر دریافت کند و یا فقط مجموعه‌ای از نقاط جذاب گردشگری را به عنوان پیشنهاد از سامانه دریافت کند. ما ستروکارونته<sup>۶۴</sup> (Console et al., ۲۰۰۳) سامانه پیشنهادگر مخصوص رانندگان است که با استفاده از رویکرد مبتنی بر دانش، جاذبه‌ها، رستوران و هتل‌ها را به کاربران پیشنهاد می‌کند. از اطلاعات بافت برای ارائه پیشنهادات مناسب به کاربران خود، مانند پیشنهادات رستوران‌هایی که در کنار پمپ‌های سوخت‌گیری قرار دارند، زمانی که میزان سوخت ماشین کم باشد، استفاده می‌کند. سامانه ارائه شده در مقاله (García-Crespo et al., ۲۰۰۹) از اطلاعات و دانش موقعیت مکانی کاربر، ترجیحات او و تاریخچه‌ای از موقعیت‌های مکانی گذشته او برای راهنمایی و ارائه پیشنهادات به او استفاده می‌کند. این سامانه، شبکه‌های اجتماعی، وب منطقی و فن آگاه از متن را برای افزایش عملکرد سامانه خود استفاده می‌کند. این سامانه راهنمای سفارشی شده ارائه می‌دهد و مشکل عدم یکپارچگی خدمات را با توجه به جستجو، پیدا کردن و ارائه کردن خدمات سفارشی شده به کاربر به کمک منطق، موقعیت جغرافیایی و فن‌های شبکه‌های اجتماعی حل کرده است. سامانه تراولر<sup>۶۵</sup> (Schiaffino and Amandi, ۲۰۰۹)، برای ارائه مجموعه‌ای گردشگری برای تعطیلات آخر هفته و سفرها طراحی شده است. این سامانه یک عامل<sup>۶۶</sup> با ترکیب رویکردهای مبتنی بر محتوا، مبتنی بر پالایشگر مشارکتی و اطلاعات دموگرافیک ساخته است. مقاله (Al-Hassan et al., ۲۰۱۵) یک پیشنهادگر ترکیبی از رویکردهای مبتنی بر پالایشگر مشارکتی و فن اندازه‌گیری منطقی، که از آنتولوژی خدمات گردشگری استرالیا برای ساخت یک ماتریس منطقی مشابهت اقلام-اقلام استفاده کرده، طراحی و ارائه

<sup>۶۳</sup> moleskiing

<sup>۶۴</sup> MASTROCARONTE

<sup>۶۵</sup> traveler

<sup>۶۶</sup> agent



کرده است. مقاله (Neidhardt et al., ۲۰۱۵) با آمیختن نقش‌های شخصیتی یاناکیز و گیبسون (پیش‌تر به آن اشاره کردیم) با دسته‌بندی روان‌شناختی پنج‌بزرگ<sup>۶۷</sup>، تعداد نقش‌ها را به هفت عدد کاهش داده‌اند و درنهایت با رویکرد جدید مبتنی بر تصویر، سعی کرده‌اند که هر کدام از نقش‌ها را با مجموعه‌ای از تصاویر بیان کنند. مقاله (Grün et al., ۲۰۱۷) در ادامه مقاله اشاره شده، سعی کرده است که هر کدام از نقش‌ها را که تشکیل‌دهنده نمایه یک کاربر هستند با ویژگی‌های مقاصد (منظور نقاط و فعالیت‌های گردشگری) منطبق کند. بدین‌صورت که کاربر پس از امتیاز دادن به هر کدام از نقش‌ها به صورت‌برداری از اعدادی خواهد بود که به نقش‌ها بنا به نزدیکی او به شخصیت خود امتیاز داده است. سپس فاصله این بردار با بردار مقاصد که توسط خبرگان بنا به نزدیکی هر کدام از نقش‌ها به آن مقصد عددی اختصاص یافته، اندازه‌گیری می‌شود و مقاصد با نزدیک‌ترین فاصله با شخصیت کاربر به کمک آنتولوژی برای او پیشنهاد می‌شود سپس بازخورد از کاربر دریافت می‌شود و نمایه کاربر به کمک آنتولوژی اصلاح می‌شود. با نگاهی که به مقالات داشتیم این نکته بسیار روشن بود که پیشنهادگرهای حوزه گردشگری از فن‌ها و رویکردهای متنوع و ترکیبی بسیار زیادی استفاده می‌کنند و در پیشنهادگرهای مبتنی بر موبایل، حس‌گرها و اطلاعات موقعیت مکانی کاربر نقش بسیار پررنگی جهت پالایش و ارائه اطلاعات و پیشنهادات مناسب به کاربر ایفا می‌کند. با مروری از ادبیات مقالات پیشنهادگر در حوزه گردشگری، درمیابیم که سامانه‌هایی که برای پیشنهاد فعالیت‌ها و مقاصد گردشگری طراحی شده‌اند به‌طور عمده از رویکردهای پالایشگر مشارکتی و رویکرد مبتنی بر محتوا استفاده کرده‌اند و در طرف دیگر ماجرا سامانه‌هایی که برای پیشنهاد یک مسیر یا یک مجموعه گردشگری طراحی شده‌اند بیشتر از رویکرد مبتنی بر دانش و بافت آگاه استفاده کرده‌اند. جدول ۲-۴ خلاصه‌ای از مقالات مرور شده در حوزه پیشنهادگرهای گردشگری است.

جدول ۲-۴ مروری بر پژوهش‌های گذشته در پیشنهادگرهای گردشگری

مرجع	هدف پژوهش	رویکرد و فن‌های مورد استفاده	ارزیابی
(Burke et al., ۱۹۹۶)	ساخت پیشنهادگر برای پیشنهاد آپارتمان‌های اجاره‌ای در شیکاگو	رویکرد مبتنی بر دانش	-
(Tung and Soo, ۲۰۰۴)	ارائه پیشنهاد رستوران در تایپه	روش بافت آگاه	آزمایش مبتنی بر سناریو (اجرای سامانه برای یک نمونه کاربر فرضی)
(Pashtan et al., ۲۰۰۳)	ارائه اطلاعات و مقاصد گردشگری	روش بافت آگاه	-

<sup>۶۷</sup> Big-five

-	رویکرد پالایشگر مشارکتی و رویکرد مبتنی بر دانش	پیشنهاد رستوران (اسپانیا)	Martinez et al., ) (۲۰۰۹)
آزمایش مبتنی بر سناریو و آزمایش مطالعه کاربر	رویکرد ترکیبی شامل رویکرد مبتنی بر محتوا، رویکرد مبتنی بر پالایشگر مشارکتی، استفاده از خوشه‌بندی و قواعد انجمنی	پیشنهادگر برای مکان‌ها و نقاط گردشگری	(Lucas et al., ۲۰۱۳)
آزمایش مبتنی بر سناریو و آزمایش مطالعه کاربر	روش‌های مبتنی بر مشارکت، مبتنی بر محتوا، بافت آگاهی و آنتولوژی و اطلاعات دموگرافیک	پیشنهادگر برای مکان‌ها و نقاط گردشگری	Moreno et al., ) (۲۰۱۳)
مطالعه کاربر، ارزیابی برون‌خط	پالایش اطلاعات بافت، آنتولوژی	پیشنهاد اطلاعات مناسب برای مکان‌های تاریخی و فرهنگی	Ruotsalo et al., ) (۲۰۱۳)
روش آزمون مبتنی بر سناریو، مطالعه کاربر	مبتنی بر پالایشگر مشارکتی	پیشنهاد نقاط جذاب تفریحی گردشگری در بستر تبادل اطلاعات کاربران	Yang and Hwang, ) (۲۰۱۳)
-	رویکرد مبتنی بر دانش	پیشنهاد فعالیت‌ها و تسهیلات گردشگری	Fesenmaier et al., ) (۲۰۰۳)
-	رویکرد مبتنی بر اعتماد به نظرات دیگران در شبکه	ارائه پیشنهاد و برنامه‌ریزی اسکی	Avesani et al., ) (۲۰۰۵)
سناریو	مبتنی بر دانش، مبتنی بر اطلاعات بافت	ارائه پیشنهادات تسهیلات برای رانندگان جاده‌ای	Console et al., ) (۲۰۰۳)
مبتنی بر سناریو	شبکه‌های اجتماعی، وب منطقی، بافت-آگاه	ارائه پیشنهادات خدمات و تورهای گردشگری	García-Crespo et ) (al., ۲۰۰۹)
مطالعه کاربران	مبتنی بر محتوا، مبتنی بر مشارکت و مبتنی بر اطلاعات دموگرافیک	یک دستیار گردشگری برای ارائه پیشنهادات و کمک به گردشگر در تصمیم‌گیری	Schiaffino and ) (Amandi, ۲۰۰۹)
مطالعه کاربران	مبتنی بر تصاویر، مبتنی بر کلیدهای شخصیتهای	پیشنهاد مکان‌های گردشگری در وین	Neidhardt et al., ) (۲۰۱۵)
مطالعه کاربران	مبتنی بر کلیدهای شخصیتهای، مبتنی بر آنتولوژی	پیشنهاد مکان‌های گردشگری در وین	(Grün et al., ۲۰۱۷)

## ۲,۵ فن‌ها و روش‌ها

همان‌طور که در بخش‌های قبل مشاهده کردیم پیشنهادگرها از فن‌ها و روش‌های متنوعی برای ارائه پیشنهاد استفاده می‌کنند که عموماً این فن‌ها و روش بر مبنای مدل‌های ریاضی و آماری است. در این بخش نگاهی خواهیم داشت به فن‌ها و روش‌هایی که ما در این پژوهش استفاده خواهیم کرد.

ممکن است ما به طور مثال علاقه‌مند به انجام یک وظیفه باشیم، بخواهیم یک پیش‌بینی دقیق بکنیم و یا یک رفتار هوشمندانه داشته باشیم. فرآیند یادگیری به طور معمول بر اساس مجموعه‌ای از مشاهدات، تجربیات مستقیم و یا دستورالعمل‌ها صورت می‌گیرد؛ بنابراین به طور کلی یادگیری ماشین در رابطه با یادگیری انجام بهتر کار یا چیزی در آینده بر اساس چیزی که در گذشته تجربه شده است. تأکید یادگیری ماشین بر خودکارسازی کردن رویکردها است. منظور، توسعه الگوریتم‌هایی است که بدون دخالت و یاری انسان فراگیرند. اغلب اوقات ما وظایف خاصی مانند پالایش داده‌های هرز، را در ذهن مدنظر داریم؛ بنابراین به جای این که کامپیوتر را به طور مستقیم برنامه‌ریزی کنیم به دنبال رویکردهایی هستیم که کامپیوتر وظایف را با برنامه‌ای که خود بر اساس مثال‌هایی که به او داده‌ایم ایجاد کرده است، انجام دهد. یادگیری ماشین یکی از زیرمجموعه‌های هوش مصنوعی است. بسیار غیرمحمول است که ما بخواهیم هرگونه سامانه هوشمند که بتواند هرکدام از امکانات که ما با سامانه‌های هوشمند مانند زبان و بینایی، ارتباط داریم را بدون یادگیری انجام دهد. همچنین ما نمی‌توانیم یک سامانه را تا زمانی که یادگیری ندارد هوشمند بدانیم، بنابراین نقطه مرکزی هوشمندی، یادگیری است (Schapire, ۲۰۰۸).

یادگیری ماشین در یک حالت کلی می‌تواند به دو گروه یادگیری با نظارت<sup>۶۹</sup> و یادگیری بی‌نظارت<sup>۷۰</sup> تقسیم‌بندی شود. یادگیری با نظارت در رابطه با نحوه ایجاد نگاشت بین مجموعه‌ای از الگوهای ورودی و برچسب‌های متناظر به هرکدام از الگوها در مجموعه داده‌ها است. زمانی که این نگاشت پیدا و مدل ساخته شد، از آن می‌توان برای پیدا کردن برچسب داده‌های ناشناس جدید استفاده کرد. یادگیری با نظارت خود نیز به دو گروه دسته‌بند<sup>۷۱</sup> و رگرسیون<sup>۷۲</sup> تقسیم می‌شود. در نوع دسته‌بند برچسب‌ها گروه‌ها گسسته (مانند داده‌های جنسیت افراد در کلاس، پسر یا دختر) و در نوع رگرسیون برچسب‌ها کمیت‌های پیوسته هستند (مانند داده‌های سهام)؛ اما در سوی دیگر، یادگیری بی‌نظارت مدل کردن ویژگی‌های داده‌ها بدون ارجاع به هرگونه برچسبی است، به عبارت دیگر گفته می‌شود اجازه دهید داده‌ها خود برای شما صحبت کنند. در اینجا نیز دو گروه داریم. در نوع اول که مو سوم به خوشه‌بندی<sup>۷۳</sup> است، داده‌ها بر

---

<sup>۶۸</sup> Machine learning

<sup>۶۹</sup> Supervised learning

<sup>۷۰</sup> Unsupervised learning

<sup>۷۱</sup> Classification

<sup>۷۲</sup> Regression

<sup>۷۳</sup> clustering

اساس شباهت ذاتی شان با یکدیگر گروه‌بندی می‌شوند اما در نوع دوم موسوم به کاهش ابعاد<sup>۷۴</sup>، الگوریتم به دنبال نمایش داده‌ها در ابعاد کمتر با حفظ خواص و ویژگی‌های مربوط به تمام داده‌ها است (VanderPlas, ۲۰۱۶).

#### تحلیل عاملی<sup>۷۵</sup>

به‌طور کلی هدف تحلیل عاملی، خلاصه‌سازی داده‌ها برای کشف الگوها و فهم بهتر داده‌ها است و این کار را با گروه کردن مجدد متغیرها در خوشه‌ها محدود بر اساس واریانس اشتراک گذاشته شده شان، است. تحلیل عاملی شامل دو نوع تحلیل عاملی اکتشافی<sup>۷۶</sup> که هدف آن کشف الگوهای پیچیده بین متغیرها و تحلیل عاملی تأییدی<sup>۷۷</sup> که هدف آن تأیید فرض‌ها است؛ که در این پژوهش، هدف ما بررسی و استفاده از تحلیل عاملی اکتشافی است.

تحلیل عاملی بر این اساس که متغیرهای قابل‌مشاهده و اندازه‌گیری، می‌توانند به متغیرهای کمتر پنهان که یک واریانس مشترک را اشتراک می‌گذارند، کاسته شوند؛ بنابراین از این جهت یک الگوریتم کاهش ابعاد محسوب می‌شود. این عامل‌ها به‌طور مستقیم قابل‌اندازه‌گیری نیستند و به کمک متغیرهای خود قابل استنباط و تفسیر هستند. در تحلیل عاملی اکتشافی فرض بر این است که  $m$  عامل مشترک پنهان وجود دارد و هدف پیدا کردن حداقل عامل مشترک است که پاسخگوی همبستگی‌ها باشد. عامل‌های مشترک آن‌هایی هستند که بیش از یک متغیر را تحت تأثیر قرار می‌دهند و عامل خاص، عاملی است که فقط یک متغیر به‌خصوص را تحت تأثیر قرار می‌دهد؛ بنابراین تحلیل عاملی برای تحقیقاتی مانند داده‌های حاصل از پرسشنامه‌ها، که متغیرهای آن‌ها بسیار زیاد است استفاده می‌شود (Yong and Pearce, ۲۰۱۳).

یونگ و پیرس از قول کومری و لی (Al and Lee, ۱۹۹۲) بیان می‌دارند برای انجام یک تحلیل عاملی، نرمال بودن یک متغیر و چند متغیره در داده‌ها باید بررسی شود. همچنین عدم وجود برون‌نهاد یک متغیره و چند متغیره در داده‌ها دارای اهمیت است. همچنین تعیین عامل بر اساس این فرض است که بین عامل‌ها و متغیرها رابطه خطی است. برای این‌که خوشه‌ای عامل حساب شود، حداقل باید دارای سه متغیر باشد البته این موضوع به طراحی و هدف مطالعه و پژوهش نیز بستگی دارد. با عامل‌هایی با دو متغیر را باید با احتیاط برخورد کرد. عامل‌هایی با دو متغیر در شرایطی مجاز هستند که آن دو متغیر دارای همبستگی بیشتر از ۰٫۷ داشته باشند. اندازه داده‌های پیشنهادی هم ۳۰۰ عدد است. هرچقدر مقدار داده‌ها افزایش یابد میزان خطا نیز کاهش می‌ابد. بار عاملی برای یک متغیر به معنا میزان مشارکت

---

<sup>۷۴</sup> Dimensionality reduction

<sup>۷۵</sup> Factorial analysis

<sup>۷۶</sup> Exploratory factorial analysis

<sup>۷۷</sup> Confirmatory factorial analysis

یک متغیر در یک عامل است. بنابراین بار عاملی بالاتر نشانگر این است که ابعاد عامل‌ها برای متغیرها پاسخگو و مناسب است. نکته بعدی، همبستگی باید ۰,۳ یا بیشتر باشد و هر چیزی کم‌تر نشانی از رابطه ضعیف بین متغیرها است. تحلیل عاملی دارای چند مولفه است که عبارت‌اند از:

**استخراج عامل‌ها:** تجزیه و تحلیل عامل بر اساس مدل عامل مشترک است که خود یک مدل نظری محسوب می‌شود. این مدل فرض می‌کند که اندازه‌گیری‌های مشاهده‌شده توسط عوامل مشترک و عوامل منحصر-به‌فرد تحت تأثیر قرار می‌گیرند و الگوهای همبستگی باید تعیین شوند. رویکردهای و فن‌های متنوعی برای استخراج عامل‌ها وجود دارد. به‌طور مثال رویکرد حداکثر احتمال<sup>۷۸</sup> تلاش می‌کند تا حداکثر احتمال نمونه-برداری ماتریس همبستگی را تحلیل کند. این رویکرد بیشتر در تحلیل عاملی تأییدی بکار می‌رود. یونگ و پیرس بنا به مقاله (Costello and Osborne, ۲۰۰۵) بیان می‌کنند رویکرد عامل اصلی محور<sup>۷۹</sup> بر این اساس است که تمام متغیرها متعلق به گروه اول است و زمانی که عامل استخراج شد ماتریس باقی‌مانده<sup>۸۰</sup> محاسبه می‌شود. همچنین بیان می‌دارند عامل‌ها به همین ترتیب استخراج می‌شوند تا زمانی که پاسخگوی میزان مناسبی از واریانس ماتریس همبستگی باشند؛ اما در طرف دیگر به نقل از (Tabachnick and Fidell, ۲۰۰۷) رویکرد تحلیل مؤلفه اصلی<sup>۸۱</sup> خود یک روش کاهش ابعاد در یادگیری ماشین است. این روش سعی می‌کند بیشترین واریانس داده‌ها را با کاهش متغیرها به کمک مجموعه‌ای از مؤلفه‌ها که در بردارنده متغیرها هستند، انجام دهد.

**روش‌های چرخش:** عامل‌ها چرخیده می‌شوند تا تفسیرپذیرتر شوند و تحلیل‌های عاملی بدون چرخش عموماً مبهم و نامفهوم هستند. هدف چرخش رسیدن به یک ساختار ساده و بهینه با عامل‌های کم است که هر کدام از متغیرها سعی می‌شود بیشترین بار عاملی<sup>۸۲</sup> را داشته باشند. درنهایت این ساختار ساده و بهینه تلاش می‌کند تا هر عامل یک خوشه‌ای از متغیرهای مرتبط باهم باشند و درنتیجه تفسیرپذیرتر خواهد شد. به‌طور مثال متغیرهایی که مرتبط با زبان هستند باید بیشترین بار عاملی در عامل مهارت‌های زبانی داشته باشند و همچنین بار عاملی آن‌ها برای مهارت‌های ریاضی باید نزدیک به صفر باشد. به‌طور کلی دو نوع چرخش متعامد<sup>۸۳</sup> و چرخش مورب<sup>۸۴</sup> داریم. در چرخش متعامد زمانی است که عامل‌ها باهم دارای زاویه ۹۰ درجه

---

<sup>۷۸</sup> Maximum likelihood

<sup>۷۹</sup> Principal axis factor

<sup>۸۰</sup> Residual matrix

<sup>۸۱</sup> Principal component analysis

<sup>۸۲</sup> Factor loading

<sup>۸۳</sup> Orthogonal rotation

<sup>۸۴</sup> Oblique rotation

هستند و به عبارتی با یکدیگر هیچ گونه ارتباط و همبستگی ندارند و در طرف مقابل چرخش مورب عامل‌ها با یکدیگر دارای ارتباط هستند. در این پژوهش از چرخش متعامد استفاده خواهد شد. دو روش مرسوم چرخش متعامد کوارتیمکس<sup>۸۵</sup> و واریمکس<sup>۸۶</sup> هستند. کوارتیمکس شامل کمینه کردن عامل‌های که برای توضیح یک متغیر استفاده می‌شود است اما در سوی دیگر واریمکس شامل کمینه کردن تعداد متغیرهایی که دارای بیشترین بار عاملی در هر عامل هستند و همچنین تلاش برای کوچک‌تر کردن بارعاملی‌های کوچک است (Yong and Pearce, ۲۰۱۳).

تفسیر بارهای عاملی: یونگ به نقل از کاستلو بیان می‌دارد هنگام تحلیل عامل‌ها نیاز است تا به بار عاملی برای تعیین میزان ارتباط توجه کنیم. عامل‌ها با کمک بزرگ‌ترین بارعاملی می‌توانند شناسایی شوند. با این وجود نیاز است تا بار عاملی‌های ضعیف و صفر را جهت تأیید بارعاملی شناسایی شده بررسی و آزمون کرد. زمانی که یک متغیر در دو یا چند عامل با بار عاملی بیش از ۰,۳۲ باشد این حالت سبب ایجاد چند عاملی می‌شود. (Costello and Osborne, ۲۰۰۵) بنا به جهت‌گیری و هدف مطالعات این متغیرهای پیچیده می‌تواند به همین صورت از این جهت که ذات این متغیر بدین گونه است باقی بماند و یا این که از پژوهش حذف گردد. راه حل دیگر استفاده از یک خط مبنا برای مشخص کردن بارعاملی‌های تأثیرگذار است. علامت بارعاملی جهت ارتباط متغیر با عامل را نشان می‌دهد و بیانگر بزرگ یا کوچک بودن بار عاملی نیست. همچنین نیاز به یک خط برش برای معنا پذیری آماری بارعاملی‌های چرخش شده است. یک قاعده کلی توجه به ارتباط بین تک‌تک عامل‌های چرخش شده و اندازه مطلق داده‌ها برداشت شده است. زمانی که اندازه داده‌ها زیاد است بارعاملی‌های کوچک‌تر اجازه دارند که در یک عامل به طور محسوس در نظر گرفته شوند. با سطح آلفا ۰,۱ برای یک مجموعه نمونه برداشت شده ۳۰۰ تایی، حداقل بار عاملی برای معنا پذیری از لحاظ آماری باید ۰,۳۲ باشد (Yong and Pearce, ۲۰۱۳).

تعداد عامل‌های ماندگار: استخراج تعداد عامل‌های کم موجب ایجاد خطا واریانس می‌شود و از طرفی دیگر استخراج عامل‌های زیاد ما را از هدف تحلیل عاملی که کاهش متغیرها و تفسیرپذیری آن‌ها بود دور می‌کند؛ بنابراین نیاز به یک معیار برای انتخاب تعداد مناسب عامل داریم و از طرف دیگر انتخاب معیار مناسب خود نیز چالش بعدی است. مقدارویژه و آزمون اسکری<sup>۸۷</sup> برای تعیین تعداد عامل‌ها ماندگار مورد استفاده قرار می‌گیرد. یک معیار سرانگشتی، معیار کیسر<sup>۸۸</sup> است. این معیار پیشنهاد می‌کند تمام عامل‌هایی که مقدار

---

<sup>۸۵</sup> Quartimax

<sup>۸۶</sup> Varimax

<sup>۸۷</sup> Scree test

<sup>۸۸</sup> Kaiser's criterion

ویژه آن‌ها بالای یک است را نگه‌داریم. معیار بعدی، معیار جولیف<sup>۸۹</sup> است که پیشنهاد می‌کند عامل‌هایی را نگه‌داریم که بالای ۰,۷ است. به‌طور معمول پیشنهاد می‌شود علاوه بر استفاده از مقدار ویژه از آزمون اسکری نیز هم‌زمان استفاده شود. این آزمون، به‌صورت یک نمودار دوبعدی است که در محور افقی تعداد عامل‌ها و در محور عمودی مقادیر ویژه مربوط است. نقطه بالای شکست نشان‌دهنده تعداد عامل‌هایی است که باید باقی بماند. آزمون اسکری زمانی قابلیت اطمینان دارد که حداقل تعداد داده‌ها ۲۰۰ و بالاتر باشد. عامل‌هایی کمتر از سه متغیر دارند، عامل‌هایی که متغیرهای بسیار زیادی دارند و یا عامل‌هایی که بار عاملی متغیرهای آن‌ها کمتر از ۰,۳۲ است نامطلوب در نظر گرفته می‌شوند.

**امتیازهای عامل:** به این معناست که یک متغیر به‌تنهایی چه امتیازی در یک عامل بدست می‌آورد. یکی از رویکردها که برای بدست آوردن امتیاز عامل، روش بارتلت<sup>۹۰</sup> است که امتیازات ناریبی که تنها با عامل خود همبستگی دارد تولید می‌کند. روش دیگر اندرسون-رابین<sup>۹۱</sup> است و امتیازاتی که تولید می‌کند بدون همبستگی و استاندارد شده است. استفاده از هر کدام از این روش‌ها بسته به هدف و رویکرد پژوهش دارد اما روش اندرسون-رابین روش معمول و پراستفاده‌تری است (Yong and Pearce, ۲۰۱۳).

یکی از مشکلات در تحلیل عامل نام‌گذاری عامل‌ها است چون ممکن است این نام بازتاب‌دهنده متغیرهای درون عامل‌ها نباشد. مشکل بعدی دشواری در تحلیل برخی از متغیرها است چون ممکن است در بیش از یک عامل قرار گیرند. ممکن است متغیرهایی باهم ارتباط داشته باشند و در یک عامل قرار بگیرند ولی هیچ رابطه معنادار و منطقی وجود نداشته باشد. در نهایت نیاز است تا پژوهشگران مطالعاتی با داده‌های زیاد در زمان مشخصی برای بررسی قابلیت اطمینان عامل‌ها انجام دهند (Sharma and Sharma, ۱۹۹۶, Yong and Pearce, ۲۰۱۳).

## ۲,۵,۲ مدل‌های دسته‌بند

همان‌طور که پیش‌تر در توضیح یادگیری ماشین گفتیم، یکی از روش‌های یادگیری ماشین، یادگیری با نظارت و نوع دسته‌بندها است. در این روش داده‌ها به دو بخش داده‌های آموزش و داده‌های آزمون تقسیم می‌شوند. ابتدا الگوریتم با کمک داده‌های آموزش نگاشت بین داده‌های متغیرهای مستقل و برجسب داده‌های متغیر وابسته را فرامی‌گیرد و سپس به کمک داده‌های آزمون، دقت و سطح عملکرد الگوریتم را

---

<sup>۸۹</sup> Jolliffe's criterion

<sup>۹۰</sup> Bartlett method

<sup>۹۱</sup> Anderson-Rubi

ارزیابی می کنند. در این بخش نگاهی به الگوریتم های دسته بند مورد استفاده در این پژوهش خواهیم داشت.

رگرسیون لجستیک<sup>۹۲</sup>: این نوع از تحلیل رگرسیونی برای زمانی که داده ها در متغیر وابسته پیوسته نیستند و به صورت دوقطبی ( صفر و یکی) هستند، مناسب است؛ مانند تمام تحلیل های رگرسیونی، یک تحلیل پیش بینی کننده است. رگرسیون لجستیک برای توصیف داده ها و تشریح رابطه بین یک متغیر وابسته با مقادیر صفر و یکی و یک یا تعدادی متغیر مستقل که می تواند مقادیر عددی، ترتیبی، اسمی، فاصله داشته باشند، استفاده می شود. به طور مثال می توان برای افراد از داده های مربوط به وزن بدنی و در صورت سیگاری بودن میزان مصرف سیگار آن ها برای تشخیص سرطانی یا غیر سرطانی بودن آن ها استفاده کنیم.

بیز ساده<sup>۹۳</sup>: بیز روشی برای دسته بندی پدیده ها، بر پایه احتمال وقوع یا عدم وقوع یک پدیده است و در نظریه احتمالات با اهمیت و پر کاربرد است. اگر برای فضای نمونه ای مفروضی بتوانیم چنان افرازی انتخاب کنیم که با دانستن این که کدام یک از پیشامدهای افراز شده رخ داده است، بخش مهمی از عدم اطمینان تقلیل میابد.

این قضیه از آن جهت مفید است که می توان از طریق آن احتمال یک پیشامد را با مشروط کردن نسبت به وقوع و یا عدم وقوع یک پیشامد دیگر محاسبه کرد. در بسیاری از حالت ها، محاسبه احتمال یک پیشامد به صورت مستقیم کاری دشوار است. با استفاده از این قضیه و مشروط کردن پیشامد مورد نظر نسبت به پیشامد دیگر، می توان احتمال مورد نظر را محاسبه کرد. (تیمورپور، بابک؛ نجفی، حیدر، ۱۳۹۴)

درخت تصمیم<sup>۹۴</sup>: درخت تصمیم یک مدل یادگیرنده است که در یک ساختار درخت مانند نمونه ها را تفکیک می کند. شکل ۲-۳ یک درخت ساده تصمیم را نمایش می دهد که یک مجموعه داده را بر اساس مقادیر دو ویژگی به سه کلاس تقسیم کرده است. درخت های تصمیم ساده برای درک نحوه دسته بندی بسیار قابل فهم هستند. نمونه های جدید بر اساس میزان پیروی از هر یک از سه شاخه موجود بر اساس ویژگی های شان دسته بندی می شوند. روش های از قبیل  $C4.5$  با یک درخت خالی شروع کرده و مرتباً داده ها را تقسیم می کنند، شاخه های درخت را می سازند تا زمانی که تمام نمونه های یک شاخه به یک دسته خاص تعلق گیرد، برگ های درختان بر اساس معیارهای خاصی ساخته می شود. میزان خطا در شاخه های درختان به اندازه کافی کم است (Han et al., ۲۰۱۱).

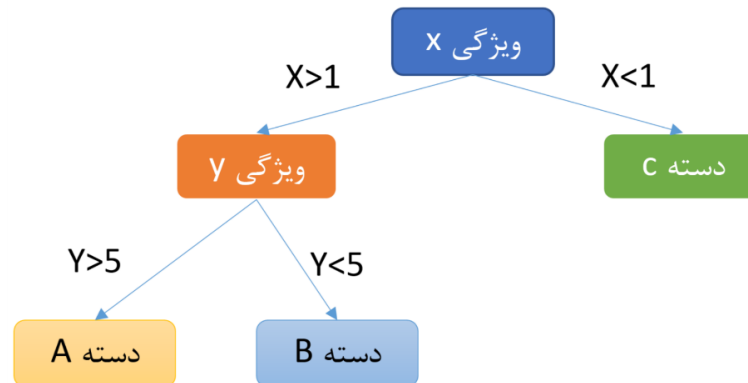
---

<sup>۹۲</sup> Logistic regression

<sup>۹۳</sup> Naïve Bayes

<sup>۹۴</sup> Decision tree





شکل ۳-۲ یک درخت ساده تصمیم (Swan et al., ۲۰۱۳)

جنگل تصادفی<sup>۹۵</sup>: جنگل تصادفی بر مبنای مجموعه‌ای از درخت‌ها که بر اساس داده‌های آموزشی مدل شده‌اند ساخته می‌شود. هر کدام از درختان تصمیم به یک زیرمجموعه از ویژگی‌های نمونه‌ها دسترسی دارند و در آخر هنگام پیش‌بینی داده‌های آزمون، هر کدام از درختان یک دسته را برای داده موردنظر پیش‌بینی می‌کنند. هر کدام از دسته‌ها که بیشترین رأی را بیاورد به‌عنوان دسته آن داده موردنظر انتخاب می‌شود (Pedregosa et al., ۲۰۱۱).

درختان اضافی<sup>۹۶</sup>: نوعی از جنگل تصادفی است که نام دیگر آن درختان به شدت تصادفی است. بر خلاف جنگل تصادفی، در هر قدم تمام مجموعه داده برای آموزش الگوریتم استفاده می‌شود و مرزهای تصمیم‌گیری به جان انتخاب بهترین به صورت تصادفی انتخاب می‌شود. معمولاً در دنیای واقعی این الگوریتم کمی از جنگل تصادفی بهتر عمل می‌کند.

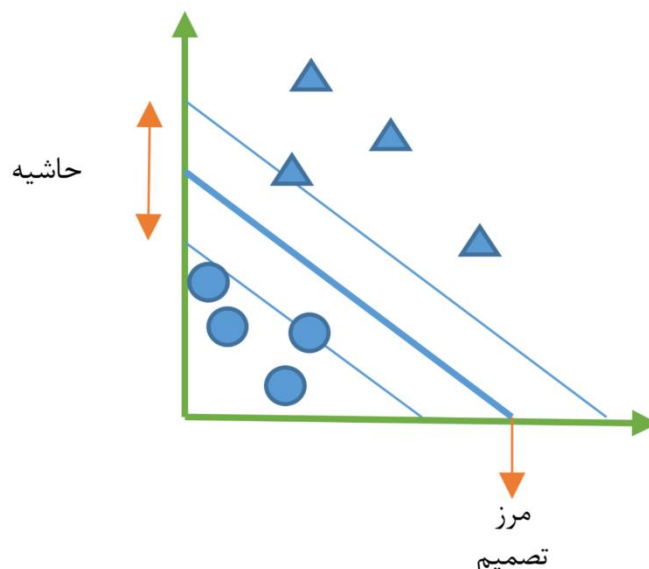
بردارهای ماشین پشتیبان یا به‌اختصار **SVM**<sup>۹۷</sup>: این مدل به‌وسیله واپنیک توسعه داده شده است. شکل ۲-۴ نحوه عملکرد الگوریتم SVM را نمایش می‌دهد. این الگوریتم بر اساس مفهوم جدا پذیر بودن خطی داده‌ها را دسته‌بندی می‌کند. ویژگی‌های به خصوصی که SVM را تعریف می‌کند عبارت‌اند از: ۱) تعیین معیاری که بهترین دسته‌بند خطی را بر اساس بیشینه کردن حاشیه‌ها تعریف کند. ۲) شناسایی بردارهای پشتیبان که کمترین تعداد نمونه داده‌های آموزشی نیاز است تا بهترین دسته‌بند خطی تعریف شود، به این

<sup>۹۵</sup> Random forest

<sup>۹۶</sup> Extra tress

<sup>۹۷</sup> Support vector machine

دلیل که آن‌ها در مرز حاشیه‌ها قرار دارند. ۳) استفاده از کرنل‌ها برای انتقال ویژگی‌های اصلی به یک فضای بزرگ‌تر غیرخطی به منظور ایجاد جدا پذیری خطی است (Han et al., ۲۰۱۱).

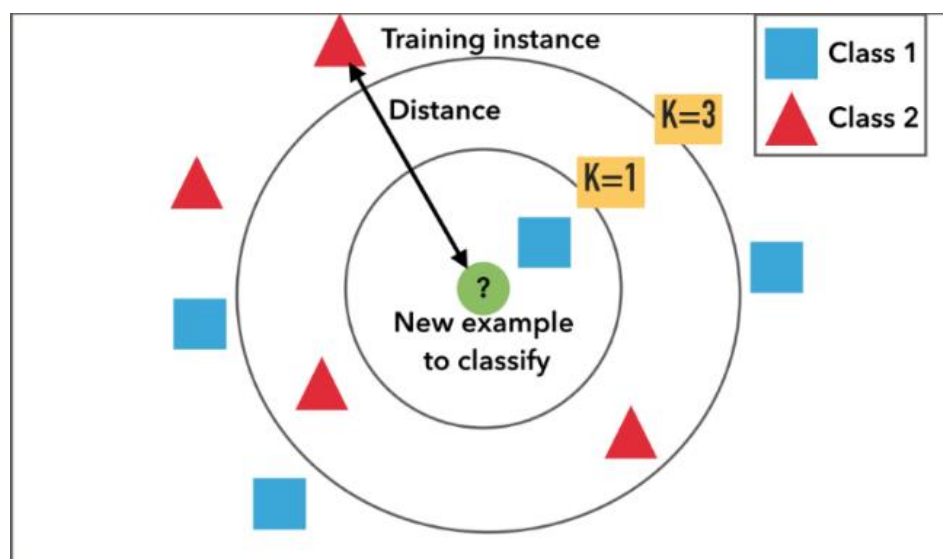


شکل ۴-۲ نحوه عملکرد الگوریتم SVM (Swan et al., ۲۰۱۳)

**الگوریتم کا-نزدیکترین همسایه**<sup>۹۸</sup>: این الگوریتم یک روش غیر پارامتریک و تنبل است. هدف آن این است که بر اساس مجموع دادگانی که به دسته‌ها مختلف دسته‌بندی شده، دسته داده‌های جدید و ناشناخته را تعیین و پیش‌بینی کند. منظور از غیر پارامتریک آن است که هیچ‌گونه فرضی در رابطه با توزیع داده‌ها انجام نمی‌دهد. به بیان دیگر ساختار مدل بر اساس خود داده‌ها شکل می‌گیرد؛ بنابراین زمانی که ما در رابطه با توزیع داده‌ها فاقد اطلاعات و یا اطلاعات کمی هستیم این الگوریتم برای استفاده گزینه مناسبی است. منظور از تنبل بودن الگوریتم این است که در فاز آموزش، به‌طور کلی آموزش صریح و خاصی صورت نمی‌گیرد و شروع به کار الگوریتم زمانی است که داده جدید وارد می‌شود سپس فاصله این داده را با سایر داده‌ها اندازه‌گیری می‌کند و برچسب نزدیک‌ترین آن‌ها را به داده جدید می‌دهد. منظور از  $k$  این است که به‌طور مثال فرض کنید در داده‌ها دو برچسب  $A, B$  داشته باشیم، اگر  $k$  را مساوی ۳ در نظر بگیریم، سه نقطه از نزدیک‌ترین فاصله‌ها را به نقطه داده جدید را انتخاب می‌کنیم و به‌طور مثال از این سه نقطه دو نقطه مربوط به برچسب  $B$  باشند و یک نقطه مربوط به برچسب  $A$  باشند آنگاه برچسب داده جدید  $B$  است. به‌طور کلی از این الگوریتم برای دسته‌بندی استفاده می‌شود ولی می‌توان برای تخمین رگرسیونی به این صورت که میانگین  $k$  نزدیک‌ترین فاصله را به عنوان پیش‌بینی رگرسیونی در نظر گرفت. مزایای این

<sup>۹۸</sup> K-nearest neighbor

الگوریتم این است که ساده، دقت بالا، هم برای دسته‌بندی و هم برای رگرسیون مناسب است و هیچ پیش فرضی در رابطه با داده‌ها ندارد و از معایب آن کند بودن الگوریتم، نیاز به حافظه بالا و همچنین به متغیرهای غیر مرتبط و داده‌ها مقیاس‌پذیر نشده حساس است. شکل ۵-۲ نمایشی تصویری از عملکرد این الگوریتم را نمایش می‌دهد (Pedregosa et al., ۲۰۱۱).



شکل ۵-۲ نمایشی از عملکرد الگوریتم کا-نزدیکترین همسایه

الگوریتم کت بوست<sup>۹۹</sup>: الگوریتم کت بوست توسط آنا ورنیکا دورگوش، واسیلی ارشوف و اندری گولین از بخش یادگیری ماشین یاندکس موتور جستجوگر روسیه‌ای ابداع شد. این الگوریتم یک روش تقویتی گرادیان است. تقویت گرادیان یک روش قدرتمند یادگیری ماشین است که نتایج قابل ملاحظه‌ای در حوزه‌های عملیاتی بدست می‌آورد. طی سالیان روش اصلی برای پرداختن به داده‌های ناهمگون، داده‌های نویزدار و شرایط پیچیده مانند جستجوگرهای وب، سامانه‌های پیشنهادگر، پیش‌بینی شرایط آب و هوایی و بسیاری از موارد دیگر بود. دلایل پشتیبانی از این الگوریتم‌ها نتایج قدرتمندی است که نشان می‌دهد چگونه دسته‌بندی‌های قدرتمند می‌توانند از تکرار ترکیب مدل‌های ضعیف‌تر با فرآیند حریصانه که مطابق با یک گرادیان نزولی در فضای عملیاتی است، ایجاد شوند. بیشتر گرادیان‌های تقویت‌شده از دسته‌بند درخت تصمیم استفاده می‌کنند. مرسوم است که از درخت تصمیم به خاطر دارا بودن ویژگی‌های عدد استفاده کنند اما تمام داده‌ها عددی نیستند و ممکن است داده‌ها طبقه‌ای داشته باشند. ویژگی‌های طبقه‌بندی ویژگی‌هایی هستند که مقادیر آن‌ها گسسته هستند و لزوماً قابل مقایسه با یکدیگر نیستند. (مانند نام

<sup>۹۹</sup> CatBoost

شهرها) یکی از متداول ترین روش ها برای پرداختن به داده های طبقه بندی شده تبدیل آن ها به داده های عددی است. الگوریتم کت بوست روش جدیدی است که برای استفاده از داده های طبقه بندی شده است و همچنین روش جدید در محاسبات برگ ها در درختان استفاده می کند که سبب می شود در مقابل بیش برآزش<sup>۱۰۰</sup> (زمانی که الگوریتم در داده های آموزش عملکرد بالایی دارد ولی در داده های آزمون عملکرد به شدت افت می کند، به عبارتی الگوریتم قابلیت تعمیم الگوها فراگرفته شده را به داده های آزمون و ناشناخته ندارد، در این حالت الگوریتم به جای یادگیری روابط و ایجاد نگاشات بین الگوهای ورودی و الگوهای خروجی با برچسب، آن ها را حفظ می کند)، قدرت بیشتری دارد (Dorogush et al).

### ۲،۵،۳ بررسی عملکرد مدل

آخرین مرحله از فرآیند یادگیری ماشین، مرحله بررسی عملکرد است. در این قسمت به چند روش برای ارزیابی عملکرد مدل های دسته بند در مواجهه با داده های آزمون می پردازیم.

دقت: دقت یک مدل بر اساس نسبت میزان داده هایی که به درستی دسته بندی شده اند به مجموع داده ها است. معیار دقت با توجه به این اگر یک دسته به خصوص به طور معناداری بیشتر از سایر دسته ها باشد می تواند همراه کننده باشد.

$$\text{رابطه ۲-۳} \quad \text{دقت} = \frac{\text{تعداد بینی پیش های درست}}{\text{مجموع بینی پیش ها}}$$

حساسیت<sup>۱۰۱</sup> و ویژگی<sup>۱۰۲</sup>: برای دودسته حاصل از مدل چهار خروجی: درست-مثبت، درست-منفی، غلط-مثبت و غلط-منفی امکان پذیر است. حساسیت نسبت تعداد نمونه های مثبت که به درستی دسته بندی شده اند به کل نمونه های مثبت است.

$$\text{رابطه ۲-۴} \quad \text{حساسیت} = \frac{\text{درست مثبت}}{\text{درست مثبت} + \text{غلط منفی}}$$

$$\text{رابطه ۲-۵} \quad \text{ویژگی} = \frac{\text{غلط مثبت}}{\text{درست منفی} + \text{غلط مثبت}}$$

امتیازاف: این امتیاز میانگین هارمونیک تو امتیاز صحت<sup>۱۰۳</sup> و بازآوری<sup>۱۰۴</sup> است؛ که به صورت زیر محاسبه می شود:

<sup>۱۰۰</sup> Over fitting

<sup>۱۰۱</sup> Sensitivity

<sup>۱۰۲</sup> Specificity

<sup>۱۰۳</sup> Precision

<sup>۱۰۴</sup> Recall

$$F1 = 2 * \frac{1}{\frac{1}{\text{بازآوری}} + \frac{1}{\text{صحت}}} \quad \text{رابطه ۶-۲}$$

$$\text{صحت} = \frac{\text{درست مثبت}}{\text{درست مثبت} + \text{غلط مثبت}} \quad \text{رابطه ۷-۲}$$

$$\text{بازآوری} = \frac{\text{درست مثبت}}{\text{درست مثبت} + \text{غلط منفی}} \quad \text{رابطه ۸-۲}$$

**منحنی ROC.** این منحنی میزان حساسیت و ویژگی یک مدل تصمیم را بر اساس تمام آستانه‌های<sup>۱۰۵</sup> ممکن تصمیم بیان می‌کند. این منحنی یک تصویر کلی از عملکرد مدل به ما می‌دهد همچنان که می‌تواند برای انتخاب بهینه سطح بهینه تصمیمات که منجر به افزایش دقت مدل دسته‌بندی می‌شود استفاده کرد. همچنان می‌توان از این منحنی برای مقایسه عملکرد چندین مدل دسته‌بندی استفاده کرد (Pedregosa et al., ۲۰۱۱, Han et al., ۲۰۱۱).

#### ۲،۵،۴ دسته‌بندی چند برچسب

در مسائل دسته‌بندی چند برچسب برخلاف دسته‌بندی‌های عادی، مجموعه دادگان می‌تواند با دو یا بیش از دو برچسب هدف متناظر باشد. می‌توان به این صورت به مسئله نگاه کرد که هر نمونه در مجموعه داده‌ها دارای ویژگی‌های است که به کمک برچسب‌ها بیان می‌شود. به‌طور مثال یک نمونه فیلم را در نظر بگیرید که می‌تواند دارای دو برچسب ژانر اجتماعی و کلاسیک باشد یا نمایه فرد در یک وبسایت گردشگری دارای برچسب‌های تاریخی، هنری، فرهنگی باشد که این برچسب‌ها در مجموع بیان‌کننده ویژگی‌های نمونه (یک فیلم یا نمایه یک شخص و ...) باشد. شکل ۶-۲ مفهوم چند برچسب را به‌خوبی نمایش می‌دهد. برای حل مسائل دسته‌بندی چند برچسب رویکردهای متنوعی وجود دارد. به‌طور کلی می‌توان روش‌ها و فن‌های مختلف را در حل این دست مسائل زیرمجموعه‌ای از رویکردهای تبدیل مسئله<sup>۱۰۶</sup>، الگوریتم‌های انطباق‌پذیر<sup>۱۰۷</sup> و یادگیری گروهی<sup>۱۰۸</sup> دانست. در رویکرد تبدیل مسئله، مسئله ما به چند مسئله دیگر تغییر پیدا می‌کند. به‌طور مثال در روش باینری رلونس<sup>۱۰۹</sup> برای هر برچسب، یک مسئله دسته‌بندی در نظر گرفته می‌شود شکل ۷-۲ مفهوم این روش را به‌خوبی نمایش می‌دهد. در روش

<sup>۱۰۵</sup> threshold

<sup>۱۰۶</sup> Problem transformation

<sup>۱۰۷</sup> Adaptive algorithm

<sup>۱۰۸</sup> Ensemble learning

<sup>۱۰۹</sup> Binary Relevance

زنجیره‌ای<sup>۱۱۰</sup> در مرتبه اول دسته‌بند سعی در نگاشت بین برچسب هدف اول و داده‌ها متناظر دارد. در مرحله بعد الگوریتم سعی در نگاشت بین برچسب هدف دوم و داده‌های متناظر به اضافه نتایج پیش‌بینی‌شده از دسته‌بند اول است و به همین ترتیب. شکل ۸-۲ مفهوم این روش را به روشنی بیان می‌کند. در روش لیبل پاورست<sup>۱۱۱</sup>، در این روش الگوریتم مسئله را به یک مسئله چند دسته و یک برچسب تبدیل می‌کند. شکل ۹-۲ به روشنی بیانگر نحوه عملکرد این الگوریتم است. الگوریتم‌های انطباق پذیر همان‌طور که از نامشان پیداست، دچار تغییراتی شده‌اند تا با مسائل چند برچسب تطبیق پیدا کنند و بدون تغییر و تبدیل در مسئله، بتوانند داده‌ها را دسته‌بندی کنند. به‌طور مثال الگوریتم کا-نزدیکتر همسایه که به‌اختصار به صورت KNN بیان می‌شود، الگوریتم انطباق‌پذیر آن برای حل مسائل چند دسته به‌اختصار MLKNN بیان می‌شود (Liu and Chen, ۲۰۱۵). اما روش‌های یادگیری گروهی، بیش از یک مدل برای مجموعه‌ای از برچسب‌ها در یک مجموعه دادگان استفاده می‌کند.

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$Y_1$	$Y_2$	$Y_3$	$Y_4$
1	0.1	3	1	0	0	1	1	0
0	0.9	1	0	1	1	0	0	0
0	0.0	1	1	0	0	1	0	0
1	0.8	2	0	1	1	0	0	1
1	0.0	2	0	1	0	0	0	1
0	0.0	3	1	1	?	?	?	?

شکل ۶-۲ مثالی از داده‌های چند برچسب (برداشت از وبسایت [www.analyticsvidhya.com](http://www.analyticsvidhya.com))

X	$Y_1$	X	$Y_2$	X	$Y_3$	X	$Y_4$
$\mathbf{x}^{(1)}$	0	$\mathbf{x}^{(1)}$	1	$\mathbf{x}^{(1)}$	1	$\mathbf{x}^{(1)}$	0
$\mathbf{x}^{(2)}$	1	$\mathbf{x}^{(2)}$	0	$\mathbf{x}^{(2)}$	0	$\mathbf{x}^{(2)}$	0
$\mathbf{x}^{(3)}$	0	$\mathbf{x}^{(3)}$	1	$\mathbf{x}^{(3)}$	0	$\mathbf{x}^{(3)}$	0
$\mathbf{x}^{(4)}$	1	$\mathbf{x}^{(4)}$	0	$\mathbf{x}^{(4)}$	0	$\mathbf{x}^{(4)}$	1
$\mathbf{x}^{(5)}$	0	$\mathbf{x}^{(5)}$	0	$\mathbf{x}^{(5)}$	0	$\mathbf{x}^{(5)}$	1


شکل ۷-۲ نمایش عملکرد روش باینری رلونس (برداشت از وبسایت [www.analyticsvidhya.com](http://www.analyticsvidhya.com))

<sup>۱۱۰</sup> Classifier Chains

<sup>۱۱۱</sup> Label powerset

X	$Y_1$	X	$Y_1$	$Y_2$	X	$Y_1$	$Y_2$	$Y_3$	X	$Y_1$	$Y_3$	$Y_3$	$Y_4$
$\mathbf{x}^{(1)}$	0	$\mathbf{x}^{(1)}$	0	1	$\mathbf{x}^{(1)}$	0	1	1	$\mathbf{x}^{(1)}$	0	1	1	0
$\mathbf{x}^{(2)}$	1	$\mathbf{x}^{(2)}$	1	0	$\mathbf{x}^{(2)}$	1	0	0	$\mathbf{x}^{(2)}$	1	0	0	0
$\mathbf{x}^{(3)}$	0	$\mathbf{x}^{(3)}$	0	1	$\mathbf{x}^{(3)}$	0	1	0	$\mathbf{x}^{(3)}$	0	1	0	0
$\mathbf{x}^{(4)}$	1	$\mathbf{x}^{(4)}$	1	0	$\mathbf{x}^{(4)}$	1	0	0	$\mathbf{x}^{(4)}$	1	0	0	1
$\mathbf{x}^{(5)}$	0	$\mathbf{x}^{(5)}$	0	0	$\mathbf{x}^{(5)}$	0	0	0	$\mathbf{x}^{(5)}$	0	0	0	1

شکل ۸-۲ نمایشی از عملکرد روش دسته بندی زنجیره ای (برداشت از وبسایت [www.analyticsvidhya.com](http://www.analyticsvidhya.com))

X	$Y_1$	$Y_2$	$Y_3$	$Y_4$		X	$Y \in 2^L$
$\mathbf{x}^{(1)}$	0	1	1	0		$\mathbf{x}^{(1)}$	0110
$\mathbf{x}^{(2)}$	1	0	0	0		$\mathbf{x}^{(2)}$	1000
$\mathbf{x}^{(3)}$	0	1	1	0		$\mathbf{x}^{(3)}$	0110
$\mathbf{x}^{(4)}$	1	0	0	1		$\mathbf{x}^{(4)}$	1001
$\mathbf{x}^{(5)}$	0	0	0	1		$\mathbf{x}^{(5)}$	0001

شکل ۹-۲ نمایشی از عملکرد روش لیبل پاورست (برداشت از وبسایت [www.analyticsvidhya.com](http://www.analyticsvidhya.com))

دسته‌بندی‌های چند برجسب کاربردهای گسترده‌ای دارند. از این نوع دسته‌بندی در علوم بیوانفورماتیک در دسته‌بندی ژن‌ها یا پیش‌بینی عملکرد چندگانه پروتئین‌ها استفاده می‌شود. همچنین برای دسته‌بندی متن‌ها، دسته‌بندی و طبقه‌بندی تصاویر و برجسب زدن به داده‌های صوتی، مانند موسیقی‌هایی که با برجسب‌های کلاسیک، جاز، آرامش‌بخش و ... در وبسایت‌ها موسیقی می‌بینیم استفاده می‌شود. همچنین بسیاری از پژوهش‌ها و بسیاری از برنامه‌های تجاری از این دسته‌بندی‌ها برای طراحی پیشنهادگرها استفاده کرده‌اند. به‌طور مثال وبسایت نتفلیکس که یک وبسایت فیلم و سریال است، برای برجسب‌گذاری و پیشنهاد محتوای خود به کاربران از مدل‌های دسته‌بندی چند برجسب استفاده می‌کند. در حوزه پژوهش به‌طور مثال مقاله (Lima and De Castro, ۲۰۱۴) از این مدل‌های دسته‌بندی برای پیش‌بینی برجسب کاربران بدون برجسب یا با برجسب کم بر اساس داده‌های افراد مرتبط با او استفاده کرده است و بررسی و دقت عملکرد الگوریتم خود را بر مجموعه داده‌های مختلف با مدل‌های متنوعی از دسته‌بندی‌های چند برجسب آزموده‌اند. مقاله (Zheng et al., ۲۰۱۴) به کمک دسته‌بندی‌های چند برجسب، برای پیش‌بینی برجسب کاربران و ارائه پیشنهاد محتوا استفاده کرده است. آن‌ها الگوریتم خود را بر روی چهار مجموعه داده متفاوت استفاده کرده‌اند. در داده‌های گرد شگری که از وبسایت تریپ ادوایزر بدست آمده است، داده‌ها شامل متغیرهایی مانند کشور کاربر، زمان در شهر کاربر، شهر هتل، زمان در شهری که هتل واقع شده، استان یا ایالت و کشور مقصد را دریافت کرده و پنج برجسب هدف که مشخص‌کننده تیپ

گردشگری شخص است را برای کاربر پیش‌بینی می‌کند. در بهترین حالت و با رویکرد تغییر مسئله به روش لیبل پاورست و استفاده از یک مدل غیرخطی SVM موسوم به SVO، به دقتی در حدود ۸۵٪ درصد در پیش‌بینی برچسب کاربران دست پیدا کرده است. مقاله (Rivoli et al., ۲۰۱۷)، برای ارائه پیشنهاد غذا به کاربران از رویکرد دسته‌بندی چند برچسب استفاده کرده است. ابتدا مجموعه داده‌ای را از طریق پرسشنامه جمع‌آوری کرده‌اند. در پرسشنامه در بخش اول از کاربر ۲۱ سؤال پرسیده شده و در بخش دوم برچسب‌های هدف (۱۲ نوع غذا) از کاربر پرسیده شده است. هدف پیش‌بینی برچسب‌ها (نوع غذاهای موردپسند کاربر) بر اساس ۲۱ سؤال و اطلاعات جمع‌آوری‌شده از آن‌ها است.

## ۲,۶ اختصارات

جدول ۲-۵ اختصارات پر کاربرد در پیشنهادگرهای گردشگری و اختصارات استفاده شده در این پژوهش

اختصار	واژه کامل	توضیحات
RS	Recommender system	سامانه‌های پیشنهادگر
POI	Point of interest	مکان یا فعالیت موردنظر گردشگری
CB	Content base	رویکرد مبتنی بر محتوا
CL	Collaborative base	رویکرد مبتنی بر مشارکت
KB	Knowledge base	رویکرد مبتنی بر دانش
SVM	Support vector machine	دسته‌بند برادرهای ماشین پشتیبان
K-NN	k-nearest neighbor	دسته‌بند ک- نزدیک‌ترین همسایه
RF	Random forest	دسته‌بند جنگل تصادفی
LR	Logistic regression	دسته‌بند رگرسیون لجستیک
DT	Decision Tree	دسته‌بند درخت تصمیم
NB	Naïve Bayes	دسته‌بند بی‌ز ساده
Ac	Accuracy	معیار ارزیابی دسته‌بند: دقت
HL	Hamming loss	معیار ارزیابی دسته‌بند: خطای همینگ
Pr	Precision	معیار ارزیابی دسته‌بند: صحت
ML/MLC	Multilabel / Multilabel classification	دسته‌بندهای چند برچسب که با ظاهر شدن در ابتدای مدل‌ها نشان از نوع تطبیق‌پذیر بود آن مدل است
BR	Binary relevance	یک رویکرد در تبدیل مسئله در دسته‌بندهای چند برچسب
LP	Label Powerset	یک رویکرد در تبدیل مسئله در دسته‌بندهای چند برچسب
CC	Classifier chains	یک رویکرد در تبدیل مسئله در دسته‌بندهای چند برچسب



## ۲,۷ خلاصه فصل و نتیجه‌گیری

فناوری اطلاعات تأثیر انکار ناشدنی بر زندگی بشر داشته است و ایجاد مزیت‌های دوسویه برای کاربران و خدمات‌دهندگان کرده است. اگرچه که در کنار این مزیت‌ها، خود زمینه‌ساز مشکلاتی مانند افزونگی اطلاعات شده است. در این میان سامانه‌های پیشنهادگر یکی از مؤثرترین راه‌حل‌ها برای برخورد با افزونگی اطلاعات است. این سامانه‌ها به‌طور عمده در سه رویکرد مبتنی بر دانش، مبتنی بر محتوا و مبتنی بر مشارکت قرار دارند. البته فن‌ها و روش‌های دیگری هم وجود دارد که در راستای این رویکردها استفاده می‌شود. پیشنهادگرها هم علاوه بر مزایا خود نیز دارای چالش‌ها و مشکلات هستند که در طول سالیان معماری و الگوریتم‌های متنوعی برای غلبه بر این چالش‌ها ارائه شده است. پیشنهادگرها به روش‌های مختلفی ارزیابی می‌شود که با بررسی در ادبیات پیشنهادگرهای گرد شگری، پرتفردارترین آن‌ها مطالعه کاربر و مبتنی بر سناریو است. در ادامه بیان کردیم یادگیری ماشین دارای دو رویکرد بی نظارت و با نظارت است که ما در این پژوهش هم از رویکرد با نظارت و هم از رویکرد بی نظارت استفاده خواهیم کرد. در رویکرد بی نظارت تحلیل عاملی، روشی است برای کاهش تعداد متغیرها و گروه‌بندی آن‌ها در عامل‌ها که موجب تفسیرپذیری داده‌ها و کاهش ابعاد مسئله می‌شود و در رویکرد با نظارت مدل‌های دسته‌بند هستند که برای پیش‌بینی برچسب هدف برای مجموعه‌ای از داده‌ها استفاده می‌شود. مدل‌های یادگیری ماشین به روش متنوع ارزیابی می‌شوند که از پرتفردارترین آن‌ها اندازه‌گیری دقت، صحبت، بازآوری و اف-امتیاز است. در مدل‌های دسته‌بند، نوعی از آن‌ها موسوم به دسته‌بندهای چند برچسب هستند که برخلاف مسائل معمول، هدف پیش‌بینی یک برچسب هدف است در اینجا هدف پیش‌بینی دو یا بیشتر برچسب هدف متناظر با داده‌ها است. این نوع از دسته‌بندها خود به سه رویکرد مختلف برای حل مسئله تقسیم‌بندی می‌شوند که شامل رویکرد تغییر و تبدیل مسئله، رویکرد الگوریتم‌های تطبیق‌پذیر و الگوریتم‌های یادگیری گروهی است که ما از هر سه دسته از این نوع دسته‌بندها در پژوهش خود استفاده خواهیم کرد.

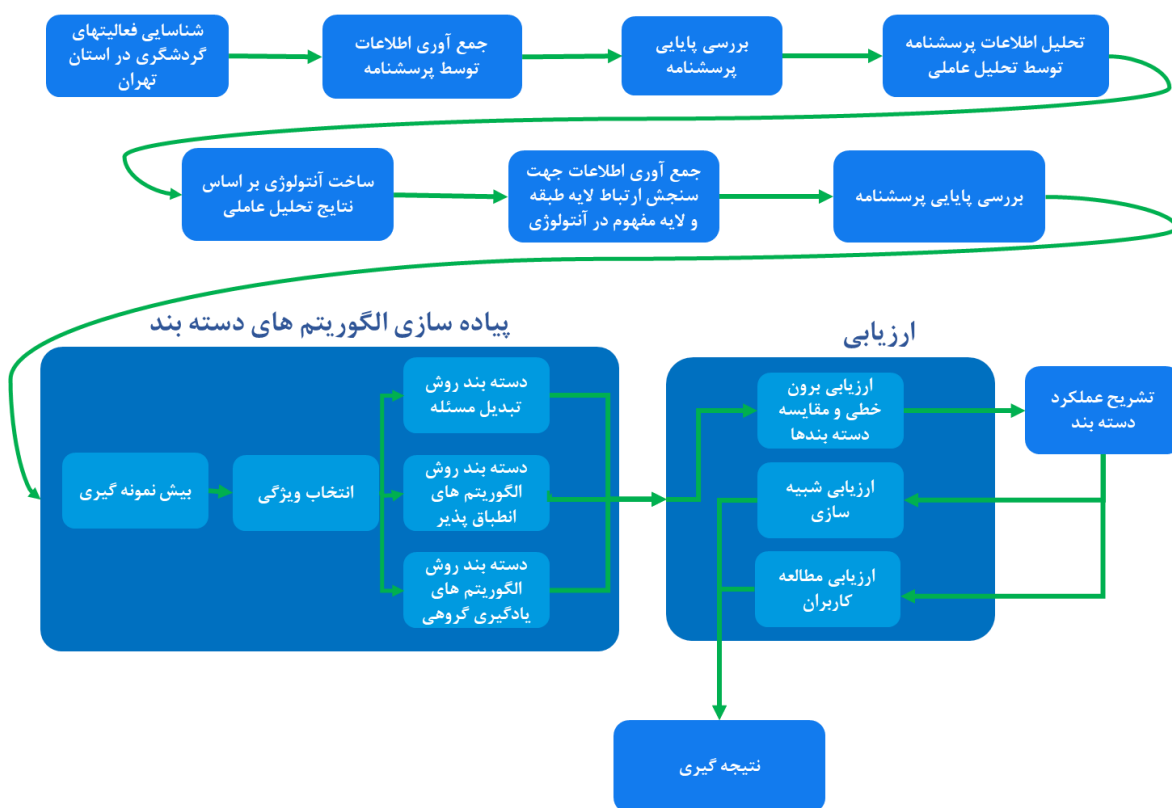
## فصل سوم

### روش شناسی پژوهش

### ۳,۱ مقدمه

در پیشنهادگرها گرد شگری که از رویکرد مبتنی بر مشارکت استفاده شده است به طور عمده از فن‌ها و روش‌های خوشه‌بندی و دسته‌بندی‌های قواعد انجمنی استفاده می‌شود؛ اما در کنار مزایا، این روش‌ها دارای مشکلاتی نیز هستند. ممکن است مجموعه‌ای از داده‌ها به طور مناسب قابل خوشه‌بندی نباشد و یا تعداد خوشه‌ها بسیار زیاد و یا بسیار کم که از ضا کننده الگوریتم پیشنهادگر نباشد. به طور معمول نمونه‌های موجود در هر کدام از خوشه‌ها، شباهت بسیاری باهم دارند و اگر صرفاً با توجه به یک خوشه پیشنهادات ارائه شود، ممکن است این پیشنهادات از تنوع مناسب برخوردار نباشد که این مشکل در ادبیات شخصی سازی بیش از حد یا به عبارت تکرار گرایی موسوم است. در پیشنهادگرهایی که از چندین خوشه برای پیشنهاد به کاربر استفاده می‌کنند، به طور معمول با توجه به خوشه‌ها، اقلام پرطرفدار هر خوشه یا متناظر با خوشه‌ها را انتخاب و به کاربر پیشنهاد می‌کنند، در این بین ممکن است اقلامی موجود باشند که در بین هیچ کدام از خوشه‌ها جزء اقلام پرطرفدار نباشد ولی افراد به خصوص به آن اقلام علاقه داشته باشند که در ادبیات به کنایه به چالش گوسفند خاکستری مشهور است. استفاده از روش‌های قواعد انجمنی نیز، زمانی که تعداد متغیرها بالا باشد و همچنین با تعداد داده‌های کم نمی‌تواند روشی اثربخش باشد به دلیل اینکه ممکن است قواعد استخراجی با توجه به بالا بودن تعداد متغیرها بسیار زیاد و باعث پیچیده‌تر شدن مسئله شود. در پیشنهادگرهای مبتنی بر مشارکت، شروع سرد مشکل رایجی است که برای حل این مشکل به طور معمول به سراغ دریافت اطلاعات صریح از کاربر می‌روند که معمولاً اطلاعات دموگرافیک از کاربران دریافت می‌شود که ممکن است برای کاربر ناخوشایند باشد و یا تعداد سوالات و یا عملگرها برای دریافت کاربر زیاد باشد که موجب کاهش رضایت و عدم همکاری کاربر شود. در پیشنهادگرهای مبتنی بر آنتولوژی بخش طراحی آنتولوژی می‌تواند چالش برانگیز و پرهزینه باشد. به دلیل این که عمدتاً ارتباط لایه‌های مختلف آنتولوژی توسط افراد خبره در حوزه‌ای که پیشنهادگر برای آن

توسعه پیدا می کند، تعیین می شود. فرض کنید تعداد مفاهیم موجود در لایه مفهوم و همچنین دسته های موجود در لایه طبقه بسیار زیاد باشد، در این صورت ارتباط این دولایه فرآیندی بسیار دشوار و زمان بر هست و همچنین نمی تواند روشی کاملاً دقیق و درست به دلیل اعمال سلیقه افراد باشد. در نتیجه ارائه روشی که به غلبه بر چالش های اشاره شده بپردازد بسیار احساس می شود. برای این منظور در این پژوهش به ارائه روشی از ترکیب مدل های دسته بندی چندرچسب و آنتولوژی خواهیم پرداخت که بتواند پاسخگوی چالش های اشاره شده باشد. در ادامه مراحل پژوهش خود و معماری الگوریتم را تشریح خواهیم کرد. شکل ۳-۱ نمای کلی از مراحل پژوهش را در یک قاب به تصویر می کشد.



شکل ۳-۱ نمای کلی از مراحل پژوهش

## ۳,۲ قدم‌ها پژوهش

در این بخش سعی می‌کنیم به‌طور مناسب به هر کدام از قدم‌هایی که در تصویر نیز به‌طور خلاصه به آن اشاره کردیم بپردازیم و دلیل استفاده از هر کدام از روش‌ها را به‌اختصار و روشنی بیان کنیم.

### ۳,۲,۱ شناسایی فعالیت‌های گردشگری

هدف از این بررسی بدست آوردن یک دسته‌بندی از مفاهیم و فعالیت‌های گردشگری در استان تهران است. استان تهران از لحاظ اهمیت سیاسی و تنوع گردشگری برای این پژوهش انتخاب شده است. شناسایی فعالیت‌ها از نقطه نظر زمانی تجربی و از نقطه نظر نحوه گردآوری و شناسایی مطالعاتی و مبتنی بر مقالات و وبسایت‌های گردشگری خواهد بود.

### ۳,۲,۲ جمع‌آوری اطلاعات پرسشنامه ۱

بعد از بدست آوردن فعالیت‌های گردشگری در استان تهران، این فعالیت‌های در قالب یک پرسشنامه که از کاربران خواسته می‌شود بنا به علاقه خود به هر کدام از فعالیت‌ها بنا به معیار لیکرت از یک تا پنج نمره دهند. پرسشنامه به صورت تصادفی از طریق شبکه‌های اجتماعی و پیام رسان در اختیار افراد قرار خواهد گرفت.

### ۳,۲,۳ بررسی پایایی<sup>۱</sup> پرسشنامه ۱

قابلیت اعتماد که واژه‌هایی مانند پایایی، ثبات و اعتبار برای آن به کار برده می‌شود، یکی از ویژگی‌های اندازه پرسشنامه و یا سایر آزمون‌ها است. مفهوم یاد شده با این امر سروکار دارد که ابزار اندازه‌گیری در شرایط یکسان تا چه اندازه نتایج یکسانی بدست می‌دهد. از روش‌های اندازه‌گیری پایایی آلفای کرونباخ<sup>۲</sup> است. این روش برای محاسبه هماهنگی درونی ابزار اندازه‌گیری از جمله پرسشنامه‌ها یا آزمون‌هایی که خصیصه‌های مختلف را اندازه‌گیری می‌کند بکار می‌رود. مقدار صفر این ضریب نشان‌دهنده عدم قابلیت اعتماد و مثبت یک نشان‌دهنده قابلیت اعتماد کامل است. فرمول زیر نحوه محاسبه آلفای کرونباخ را برای داده‌های طیف لیکرت نمایش می‌دهد.

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum_{i=1}^k \sigma_{Y_i}^2}{\sigma_x^2} \right)$$

<sup>۱</sup> Reliability

<sup>۲</sup> Cronbach's alpha

#### ۳,۲,۴ تحلیل عاملی

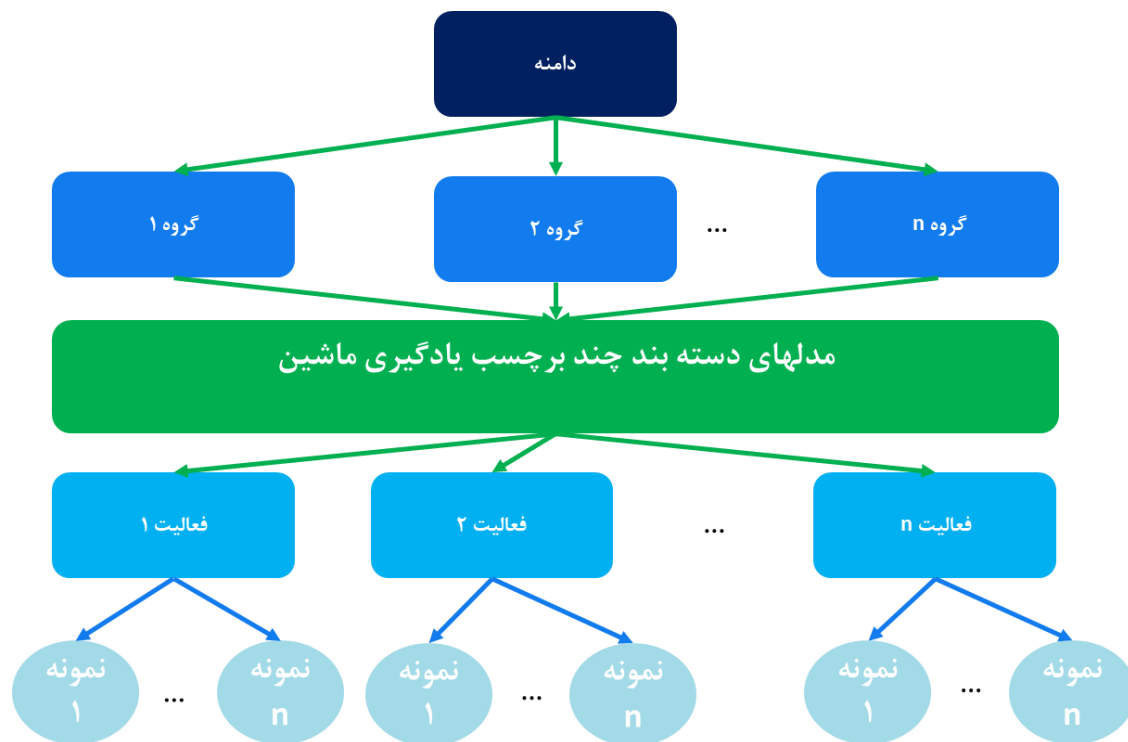
بعد از بررسی پایایی پرسشنامه، انجام تحلیل عاملی بر روی داده‌های حاصل از پرسشنامه است. هدف از انجام این کار بدست آوردن گروه‌های مناسب از فعالیت‌های گردشگری و ساخت کلیشه‌های گردشگری است.

#### ۳,۲,۵ ساخت آنتولوژی

آنتولوژی موردنظر دارای چهار سطح خواهد بود که سطح اول نام دامنه که همان گردشگری است. سطح بعد گروه‌ها است که عناصرا این سطح نتایج تحلیل عاملی و عامل‌های استخراج شده خواهد بود. سطح بعدی، سطح مفهوم خواهد بود که شامل فعالیت‌ها شناسایی شده گردشگری در استان است که در این پژوهش ما به آن برچسب‌های نمایه کاربر نیز می‌گوییم و در سطح آخر یا سطح نمونه، نمونه‌هایی از هرکدام از فعالیت‌های سطح قبلی در آن خواهد بود. هدف کشف ارتباط سطح گروه‌ها با سطح مفاهیم به کمک یادگیری ماشین و استفاده از رویکرد یادگیری با نظارت و مدل‌های چندبرچسب است. شکل ۲-۳ نمایی از موارد تشریح شده را به‌روشنی بیان می‌کند.

#### ۳,۲,۶ جمع‌آوری اطلاعات پرسشنامه ۲

این بار هدف از جمع‌آوری پرسشنامه، سنجش ارتباط بین سطح گروه که از نتایج تحلیل عاملی بدست آمده با لایه مفهوم که شامل فعالیت‌های شناسایی شده است. این پرسشنامه دارای دو بخش خواهد بود. در بخش اول نظر کاربر نسبت به هرکدام از گروه‌ها سطح اول آنتولوژی با معیار لیکرت سنجیده خواهد شد. در بخش دوم از کاربران خواسته می‌شود فعالیت‌ها موردعلاقه خود را از لیست ارائه‌شده به آن‌ها انتخاب کنند. سپس داده‌های حاصل از این پرسشنامه یک مجموعه داده برچسب‌دار برای انجام دسته‌بندی فراهم خواهد کرد. پرسشنامه به صورت تصادفی از طریق شبکه‌های اجتماعی و پیام‌رسان در اختیار افراد قرار خواهد گرفت.



شکل ۲-۳ نمایی از آنتولوژی مورد نظر پژوهش

### ۳،۲،۷ بررسی پایایی پرسشنامه ۲

به همان شکلی که پیش تر گفته شد پایایی این پرسشنامه نیز توسط آلفای کرونباخ بررسی خواهد شد.

### ۳،۲،۸ پیاده سازی الگوریتم های دسته بند چند برچسب

برای پیاده سازی الگوریتم های یادگیری ماشین معمولاً نیاز است داده ها پیش پردازش شوند. پیش پردازش شامل پاک سازی داده ها از داده های بی معنی، اشتباه و فاقد اطلاعات است و کاهش داده ها که یکی از روش ها آن کاهش ابعاد و انتخاب ویژگی<sup>۳</sup> است به منظور انجام سریع تر محاسبات است (Ren, ۲۰۱۳). ما در این پژوهش از انتخاب ویژگی به طور غیرمستقیم استفاده خواهیم کرد. منظور از غیرمستقیم، این است که بسیاری از الگوریتم های توسعه داده شده در پکیج های یادگیری ماشین به طور خودکار قابلیت انتخاب ویژگی دارند. همچنین داده ها را به داده های نرمال تبدیل خواهیم کرد. مسئله دیگری که در پیش پردازش داده ها به آن خواهیم پرداخت نامتوازن بودن داده ها است. منظور از نامتوازن بودن داده ها، این است که داده های مربوط به یک کلاس به طور معنی داری بسیار بیشتر از کلاس دیگر باشد؛ مثلاً در داده های بانکی افرادی که دارای معوقه در اقساط خود هستند در مقابل سایر افراد بسیار کم باشد بنابراین امکان یادگیری

<sup>۳</sup> Feature selection

و محاسبه دقت الگوریتم به شدت کاهش پیدا می کند. فرض کنید مجموعه دادگانی داریم که کلاس های هدف برای آن "بله" و "خیر" است. تعداد داده های مربوط به "بله" ۹۹ باشد و تعداد داده های مربوط به "خیر" یک باشد. اگر الگوریتم تمام صد داده را "بله" پیش بینی کند، دقت الگوریتم ۹۹٪ خواهد بود که یک نتیجه کاملاً گمراه کننده است؛ بنابراین صحت و بازآوری الگوریتم که قبلاً نیز بحث شد سنجیده می شود. راه برخورد با مشکل داده های نامتوازن استفاده از بیش نمونه گیری<sup>۴</sup> است. بیش نمونه گیری دارای روش ها متنوعی است. یکی از روش ها تکرار تصادفی مجموعه دادگانی که کلاس آن ها در اقلیت است. روشی که ما در این پژوهش افتاده خواهیم کرد موسوم به فن بیش نمونه گیری اقلیت مصنوعی یا به اختصار SMOTE<sup>۵</sup> است. در این روش به جای تکرار داده های که کلاس آن ها در اقلیت است که خود منجر به بیش برآزش ممکن است شود، داده ها به طور مصنوعی تولید می شوند. روش کار بدین صورت است که یک نقطه اقلیت به تصادف انتخاب می شود و کا-نزدیکترین هم سایه آن را انتخاب می کنیم و از نقطه فعلی به کا نزدیک ترین هم سایه یک خط رسم می کنیم و از آن یک بردار رسم می کنیم. سپس این بردار را دیک عدد تصادفی بین صفر و یک ضرب می کنیم و نقطه جدید بدست می آید (Gosain and Sardana, ۲۰۱۷). برای پیش بین سطح مفهوم که فعالیت های گرد شگری در آن سطح قرار دارند به وسیله داده های سطح گروه که گروه های ایجاد شده از تحلیل عاملی در آن قرار دارند از الگوریتم های دسته بند چند برچسب استفاده خواهیم کرد. به عبارتی با کمک امتیازی که کاربر به هر کدام از گروه های می دهد علاقه او به فعالیت های گرد شگری را پیش بینی خواهیم کرد. بدین منظور از سه رویکرد دسته بند های چند برچسب که در فصل قبل تشریح شد استفاده خواهیم کرد. در رویکرد تبدیل مسئله و روش لیبیل پاورست و روش دسته بندی زنجیره ای از الگوریتم های naive Bayes, SVM, random forest, k-NN, decision Tree, linear logistic استفاده خواهیم کرد و در روش باینری رلونس علاوه بر الگوریتم های اشاره شده از catboost نیز استفاده خواهیم کرد. در رویکرد الگوریتم های تطبیق پذیر از BRKNNaClassifier, MLKNN که الگوریتم های تطبیق پذیر شده از الگوریتم KNN (کا-همسایه) است، استفاده خواهد شد. در رویکرد الگوریتم های یادگیرنده گروهی نیز از الگوریتم های جنگل تصادفی و درختان اضافی استفاده خواهد شد.

### ۳،۲،۹ ارزیابی برون خطی

در بخش ارزیابی، اول به سراغ ارزیابی برون خطی جهت سنجش عملکرد الگوریتم های اشاره شده خواهیم رفت. عملکرد این الگوریتم ها در پیش بینی فعالیت ها (برچسب های نمایه کاربر) برای هر کاربر به طور کلی

<sup>۴</sup> oversampling

<sup>۵</sup> Synthetic oversampling technique



نمایانگر عملکرد پیشنهادگر ما خواهد بود. محاسبات ارزیابی دسته‌بندهای چند برجسب با دسته‌بندهای معمولی تفاوت‌هایی دارد. در این پژوهش قصد داریم تا میزان دقت، صحت، بازآوری و تابع زیان همینگ<sup>۶</sup> را استفاده کنیم. در فصل دوم با مفهوم سه معیار اول آشنا شدیم در ادامه نحوه محاسبه آن را بیان می‌کنیم.

دقت یا امتیاز همینگ<sup>۷</sup>: در محاسبه دقت یا امتیاز همینگ همچنان مانند روش‌های قبل دقت، تعداد پیش‌بینی‌های درست را بر تعداد کل پیش‌بینی‌ها تقسیم می‌کنیم.

$$\text{رابطه ۱-۳} \quad \text{دقت} = \frac{\text{تعداد پیش‌بینی‌های درست}}{\text{تعداد کل پیش‌بینی‌ها}}$$

صحت و بازآوری: اما در اندازه‌گیری صحت و بازآوری دو رویکرد کلان<sup>۸</sup> و کوچک<sup>۹</sup> وجود دارد. در رویکرد کلان صحت و بازآوری را برای هر کدام از دسته‌ها اندازه‌گیری می‌کند و سپس میانگین آن‌ها را می‌گیرد اما در رویکرد کوچک "درست-مثبت"، "غلط-مثبت" و "غلط-منفی" یا به اختصار به ترتیب NF, FP, TP تمام دسته‌ها را شمارش می‌کند و صحت و بازآوری را محاسبه می‌کند. برای اندازه‌گیری عملکرد کلی الگوریتم‌ها از رویکرد کلان استفاده می‌شود اما برای داده‌هایی که متوازن نباشند مناسب نیست و رویکرد کوچک روشی سخت‌گیرانه‌تر و برای داده‌های نامتوازن نیز مناسب است. با توجه به نامتوازن بودن داده‌های ما، از رویکرد کوچک استفاده خواهیم کرد. از بیش نمونه‌گیری فقط برای الگوریتم کت‌بوست استفاده خواهیم کرد و برای سایر الگوریتم‌ها از امکانات خود آن‌ها برای برخورد با داده‌های نامتوازن استفاده خواهد شد. اما در این میان در الگوریتم‌های تطبیق‌پذیر امکان متوازن سازی داده‌ها وجود ندارد و الگوریتم‌های کا-نزدیکترین همسایه و بیز ساده مشکلی در پردازش داده‌های نامتوازن ندارند. به همین دلیل برای سنجش دقیق عملکرد الگوریتم‌ها از معیار دقت و چهار معیار دیگری که نام خواهیم برد استفاده خواهیم کرد.

$$\text{رابطه ۲-۳} \quad \text{بازآوری} = \frac{\text{جمع تک‌تک صحت‌های محاسبه شده برای هر کدام از دسته‌ها}}{\text{جمع تعداد دسته‌ها}}$$

$$\text{رابطه ۳-۳} \quad \text{بازآوری} = \frac{\text{جمع تک‌تک بازآوری‌های محاسبه شده برای هر کدام از دسته‌ها}}{\text{جمع تعداد دسته‌ها}}$$

<sup>۶</sup> Hamming loss

<sup>۷</sup> Hamming score

<sup>۸</sup> macro

<sup>۹</sup> micro

امتیاز اف: این امتیاز، میانگین هارمونیک امتیازهای صحت و بازآوری است.

$$\text{رابطه ۳-۴} \quad \text{امتیاز اف} = \frac{(\text{بازآوری} * \text{صحت} * ۲)}{(\text{بازآوری} + \text{صحت})}$$

تابع خطا همینگ در یک بیان ساده، تابع خطا همینگ نسبت برچسب‌هایی که به درستی پیش‌بینی نشده‌اند را نمایش می‌دهد. نحوه محاسبه تابع خطا همینگ به صورت زیر است:

$$\text{رابطه ۳-۵} \quad HL = \frac{1}{|N| \cdot |L|} \sum_{i=1}^{|N|} \sum_{j=1}^{|L|} xor(y_{i,j}, z_{i,j})$$

N بیانگر تعداد نمونه‌ها در مجموعه داده است، L بیانگر تعداد برچسب‌ها در مجموعه داده است. y نشان‌دهنده برچسب هدف که نمایانگر مقادیر درست است و z نشان‌دهنده مقادیر پیش‌بینی شده است (Nooney, ۲۰۱۸).

#### ۳,۲,۱۰ تشریح عملکرد دسته‌بند

بعد از ارزیابی برون خطی، به تشریح نحوه عملکرد دسته‌بند و نحوه ارتباط بین عناصر سطح گروه و عناصر سطح مفهوم (فعالیت‌ها یا همان برچسب نمایه کاربر) خواهیم پرداخت. بدین منظور میزان اثرگذاری هر کدام از عناصر سطح گروه در آنتولوژی برای پیش‌بینی فعالیت‌های سطح مفهوم را بررسی خواهیم کرد.

#### ۳,۲,۱۱ ارزیابی مبتنی بر سناریو

برای درک و فهم نحوه کارکرد سامانه پیشنهادگر، کاربری فرضی را در نظر گرفته و مراحل عملکرد سامانه را با توجه به کاربر و داده‌های کاربر بررسی خواهیم کرد.

#### ۳,۲,۱۲ ارزیابی مطالعه کاربر

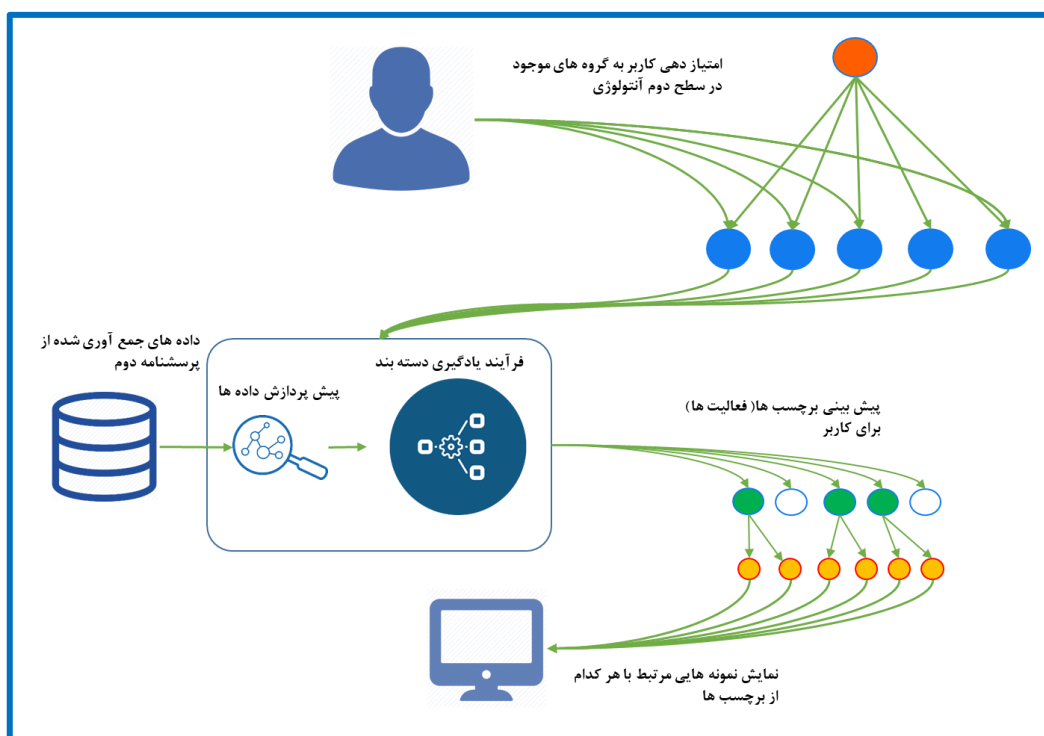
در آخرین نوع ارزیابی، عملکرد سامانه را در دنیای واقعی خواهیم سنجید. به این صورت که از تعدادی کاربر خواسته می‌شود در قالب پرسشنامه اطلاعاتی را ارائه دهند. پس از وارد کردن اطلاعات این افراد به سامانه و پیش‌بینی برچسب‌ها (فعالیت‌های موردعلاقه افراد) برای افراد، سپس این برچسب‌های پیش‌بینی‌شده همراه با نمونه عینی فعالیت‌های مرتبط با هر برچسب به کاربر ارائه می‌شود و از کاربر مجدداً درخواست می‌شود که به سؤالات پرسشنامه جهت سنجش میزان رضایت او از پیش‌بینی‌های پیشنهادگر پاسخ دهد.

#### ۳,۲,۱۳ نتیجه‌گیری

آخرین مرحله پژوهش به نتیجه‌گیری پژوهش و بررسی ارتباط ارزیابی برون خطی و مطالعه کاربر خواهیم پرداخت.

### ۳,۳ خلاصه و جمع‌بندی فصل: تشریح کلی عملکرد سامانه

همانگونه که در شکل ۳-۳ مشاهده می‌کند ابتدا با مجموعه داده‌های جمع‌آوری شده توسط پرسشنامه دوم مدل آموزش می‌بیند تا رابطه بین گروه‌های سطح اول آنتولوژی و لایه مفهوم که فعالیت‌ها یا همان برچسب نمایه کاربر است را کشف کند. سپس کاربر با ورود به سامانه و امتیاز دادن به گروه‌های سطح دوم آنتولوژی، اطلاعات وارد مدل دسته‌بندی چندبرچسب شده و فعالیت‌های کاربر (برچسب) را پیش‌بینی می‌کند و متناسب با هر کدام از فعالیت‌ها نمونه‌های مرتبط با آن فعالیت را نمایش می‌دهد.



شکل ۳-۳ تصویر کلی از عملکرد سامانه پیشنهادگر

## **فصل چهارم**

### **تجزیه و تحلیل یافته های پژوهش**

## ۴،۱ مقدمه

در فصل گذشته در ابتدا ضعف‌های پیشنهادگرها مبتنی بر مشارکت را شرح دادیم و دلایل خود برای پرداختن به این موضوع ارائه دادیم. سپس یک نقشه‌راه برای طراحی سامانه گردشگری خود ارائه داده و رویکرد خود را تشریح کردیم. در این فصل در ادامه فصل گذشته، به پیاده‌سازی و اجرای مراحل اشاره شده خواهیم پرداخت و نتایج آن را گزارش خواهیم کرد. مورد مطالعاتی ما، استان تهران بدلیل اهمیت ژئوپولیتیکی و تنوع فعالیت‌های گردشگری در آن است. بنابراین در قسمت‌های پیش‌رو، فعالیت‌ها و مکان‌های گردشگری که شناسایی و برای تحلیل استفاده شده مربوط به استان تهران است.

## ۴،۲ شناسایی فعالیت‌های گردشگری

برای شناسایی فعالیت‌های گردشگری در استان تهران، به روش کتابخانه‌ای عمل کردیم. ابتدا بخشی از فعالیت‌های گردشگری را از طریق مقالات بدست آوردیم اما این روش به طور کامل پاسخگوی نیاز ما نبود و دارای نواقصی بود. فعالیت‌های گردشگری بسیاری از مقالات مربوط و مختص به همان شهرها بود و مصداقی در استان تهران نداشت و از طرفی دیگر امکان انجام فعالیت‌های گردشگری و تفریحی بخصوصی نیز در تهران وجود دارد که در سایر شهرها و کشورها مصداق ندارد. بنابراین علاوه بر مقالات، به مطالعه و بررسی وبسایت‌های گردشگری و برنامه‌های تورهای گردشگری در استان تهران پرداختیم. در مجموع ۲۶ فعالیت گردشگری در استان تهران را شناسایی کردیم. جدول ۱-۴ لیستی از فعالیت‌های شناسایی شده و همچنین معادل/اختصار آن‌ها به زبان انگلیسی را نمایش می‌دهد. دلیل استفاده از معادل/اختصار انگلیسی، عدم سازگاری بسیاری از برنامه‌ها مورد استفاده در پژوهش با زبان فارسی بود و در گزارشات این نرم افزارها که در قسمت‌های پیش‌رو به آن‌ها خواهیم پرداخت از همین معادل/اختصارات استفاده شده است.

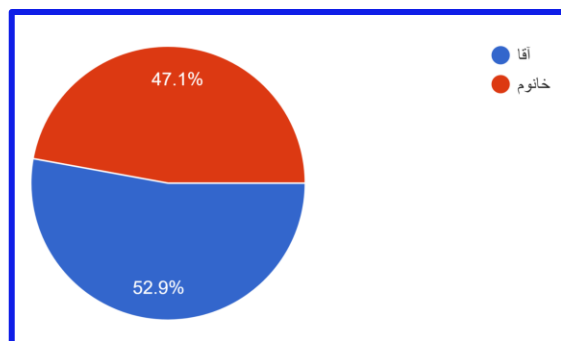
جدول ۴-۱ فهرستی از فعالیت های شناسایی شده و معادل/اختصار آن ها به زبان انگلیسی

معادل/اختصارات	نمونه‌ای از نقطه مورد نظر برای فعالیت	فعالیت
Palace	کاخ مرم، کاخ سعادت‌آباد و ..	کاخ‌گردی
TraMa	بازار تجریش و ...	بازارچه‌سنتی گردی
Musem	موزه ملی، موزه هنرهای معاصر و ...	موزه‌گردی
Mansion	باغ فردوس، عمارت کلاه فرنگی و ...	بازدید از عمارت‌ها و خانه‌های مشاهیر
Cinema	سینما کوروش و ...	رفتن به سینما
Theater	تئاتر شهر، تالار مولوی و ...	رفتن به تئاتر
CultureMusem	موزه فرش، موزه قرآن کریم و ...	بازدید از موزه‌ها فرهنگی
EmamZadeh	اما زاده صالح، امام زاده داود و ...	زیارت امام‌زاده‌ها
Mosque	مسجد ارگ، مسجد شاه و ...	حضور در مسجد
Tombs	قبرستان ظهیرالدوله، ابن بابویه و ...	زیارت قبور مشاهیر
SportEvents	مسابقات فوتبال، والیبال و ...	حضور در رویدادهای ورزشی
WaterSport	شنا، قایق سواری و ...	انجام ورزش‌های آبی
NoneWaterSport	فوتبال، والیبال و ...	انجام ورزش‌های غیر آبی
BookExib	گالری نقاشی، نگارخانه‌ها و ...	بازدید نمایشگاه‌های کتاب
ArtExib	باغ کتاب و ...	بازدید نمایشگاه‌های هنری
MusicCon	موسیقی سنتی، کلاسیک و ...	رفتن به کنسرت موسیقی
Mall	پالادیوم و ...	پاساژ گردی و خرید
Garden	باغ پرندگان، باغ نگارستان و ...	گشت و گذار در باغ‌های عمومی
Zoo	باغ وحش ارم و ...	بازدید از باغ وحش
Restaurant	ریحون، زرچ و ...	رستوران رفتن
Café	کافه نادری و ...	کافه گردی
Parks	بوستان آب و آتش، بوستان نهج‌البلاغه و ...	تفریح کردن در بوستان‌ها
Rural	روستای برغان و ...	روستا گردی
Lakes	چینگر و ...	بازدید از دریاچه‌ها
WaterfallsAndRivers	آبشار اوسند، رودخانه کرج و ...	بازدید از رودخانه‌ها و آبشارها
Mountains	درکه، توچال و ...	کوهنوردی و پیاده‌روی

## ۴,۳ جمع‌آوری اطلاعات (پرسشنامه ۱)

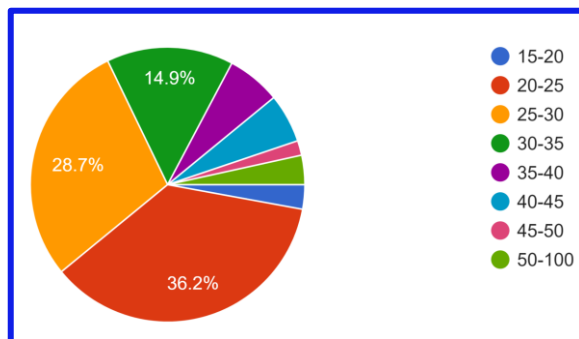
برای ایجاد گروه‌های مناسب از ۲۶ فعالیت‌های گردشگری که این گروه‌ها تشکیل دهنده سطح دوم آنتولوژی هستند نیاز به داده بود. بنابراین این ۲۶ فعالیت را در قالب یک پرسشنامه به صورت تصادفی در اختیار افراد در قرار دادیم و از آن‌ها خواستیم طبق معیار لیکرت از یک تا پنج که نشان دهنده میزان

علاقه آن‌ها به هر کدام از فعالیت‌ها ارائه شده است نمره دهند. فرم پرسشنامه در پیوست ((الف)) آمده است. در کنار پرسش در رابطه با فعالیت‌ها، جنسیت افراد و بازه سنی آن‌ها را جویا شدیم تا از جمع‌آوری متوازن پرسشنامه اطمینان حاصل کنیم. پرسشنامه به صورت الکترونیکی تهیه و از طریق شبکه‌های مجازی و پیام‌رسان‌ها در اختیار افراد قرار داده شد. به طور کلی از تاریخ اواخر بهمن ۱۳۹۶ تا اوایل اردیبهشت ۱۳۹۷، ۲۷۲ پرسشنامه جمع‌آوری شد. همانگونه که در نمودار زیر مشاهده می‌کنید نسبت جنسیتی افراد که سوالات پرسشنامه را پاسخ داده‌اند تقریباً برابر است و اختلاف ناچیز است.



شکل ۱-۴ نمودار نسبت جنسیتی پاسخ دهندگان به پرسشنامه

همچنین با مشاهده نمودار زیر در میابیم، بیشترین افرادی که به پرسشنامه پاسخ داده‌اند در بازه‌های سنی ۲۰ تا ۳۵ سال قرار دارند و البته این نتیجه به دلیل دسترسی و علاقه این بازه سنی از افراد به اینترنت و فناوری‌های مربوط به آن، کاملاً قابل پیش‌بینی و منطقی بود.



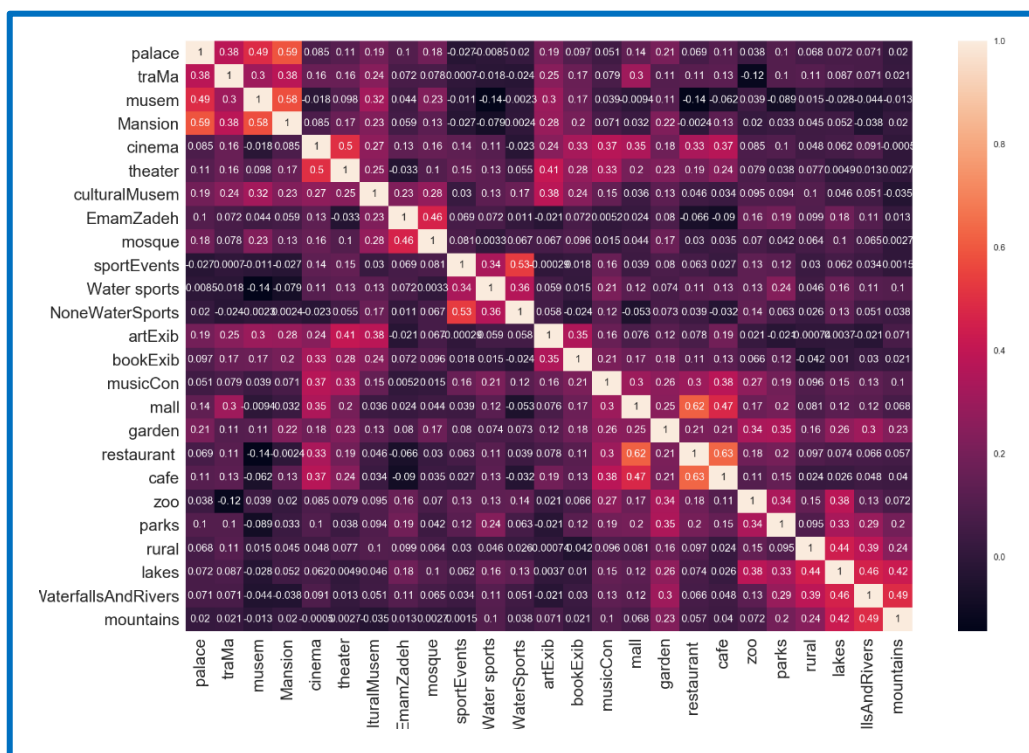
شکل ۲-۴ نمودار نمایش درصدی بازه سنی افراد به پرسشنامه اول

## بررسی پایایی پرسشنامه

برای بررسی آلفای کرونباخ، جهت سنجش پایا بودن پرسشنامه، از نرم افزار اسپس اس<sup>۱</sup> نسخه ۲۲ استفاده کردیم. میزان آلفای کرونباخ برای پرسشنامه حاضر با ۲۶ متغیر (۲۶ فعالیت سوال شده) برابر ۰,۸۴۶ است که به طور چشمگیری بیش تر از حداقل امتیاز لازم برای پایایی پرسشنامه (۰,۷) بیشتر است و در نتیجه پرسشنامه ما پایا است.

## ۴,۴ تحلیل عاملی پرسشنامه

نتایج پرسشنامه برای یافتن گروه ها و ارتباط بین فعالیت ها مورد تحلیل عاملی قرار گرفت. تصویر زیر نمودار گرمایی همبستگی پیرسون بین فعالیت ها را نمایش می دهد که به وسیله بسته سی بورن<sup>۲</sup> در زبان برنامه نویسی پایتون مصور سازی شده است.



شکل ۳-۳ نمودار گرمایی همبستگی فعالیت های گردشگری شناسایی شده

<sup>۱</sup> SPSS-۲۲

<sup>۲</sup> seaborn



همانطور که ما انتظار داشتیم، بین بسیاری از فعالیت‌های همبستگی مناسبی برقرار بود. مثلاً کاملاً قابل انتظار است افرادی که به فعالیت موزه‌گردی امتیاز بالایی داده‌اند به فعالیت کاخ‌گردی هم امتیاز بالایی بدهند. می‌شود اینطور استنباط کرد که این دسته افراد به تاریخ علاقه‌مند باشند. برای روشن شدن دقیق‌تر مسئله داده‌ها را با نرم‌افزار اسپاس نسخه ۲۲ مورد تحلیل عاملی قرار دادیم. برای استخراج عامل‌ها از رویکرد عامل اصلی محور که برای تحلیل عاملی اکتشافی مناسب است استفاده کردیم. برای چرخش، از روش چرخش واریمکس با تکرار ۳۰ استفاده کردیم. برای بررسی بسنده و مناسب بودن نمونه‌گیری به سراغ دو شاخص کیسِر و بارتلت رفتیم. حداقل مقدار شاخص کیسِر برای تحلیل عاملی مقدار ۰,۵ است که مقدار این شاخص برای پژوهش ما برابر با ۰,۷۶ است. آزمون بارتلت یک آزمون فرض آماری می‌گیرد و فرض صفر آن بیان می‌کند که ماتریس همبستگی یک ماتریس همبستگی است بنابراین ارتباط معنا داری بین متغیرها وجود ندارد. همانطور که می‌بینیم در جدول زیر مقدار آماره آزمون در ناحیه رد قرار نمی‌گیرد. (مقدار sig باید کمتر از ۰,۰۵ باشد که برای دادگان ما صفر است).

جدول ۲-۴ نتیجه آزمون بارتلت و کیسِر

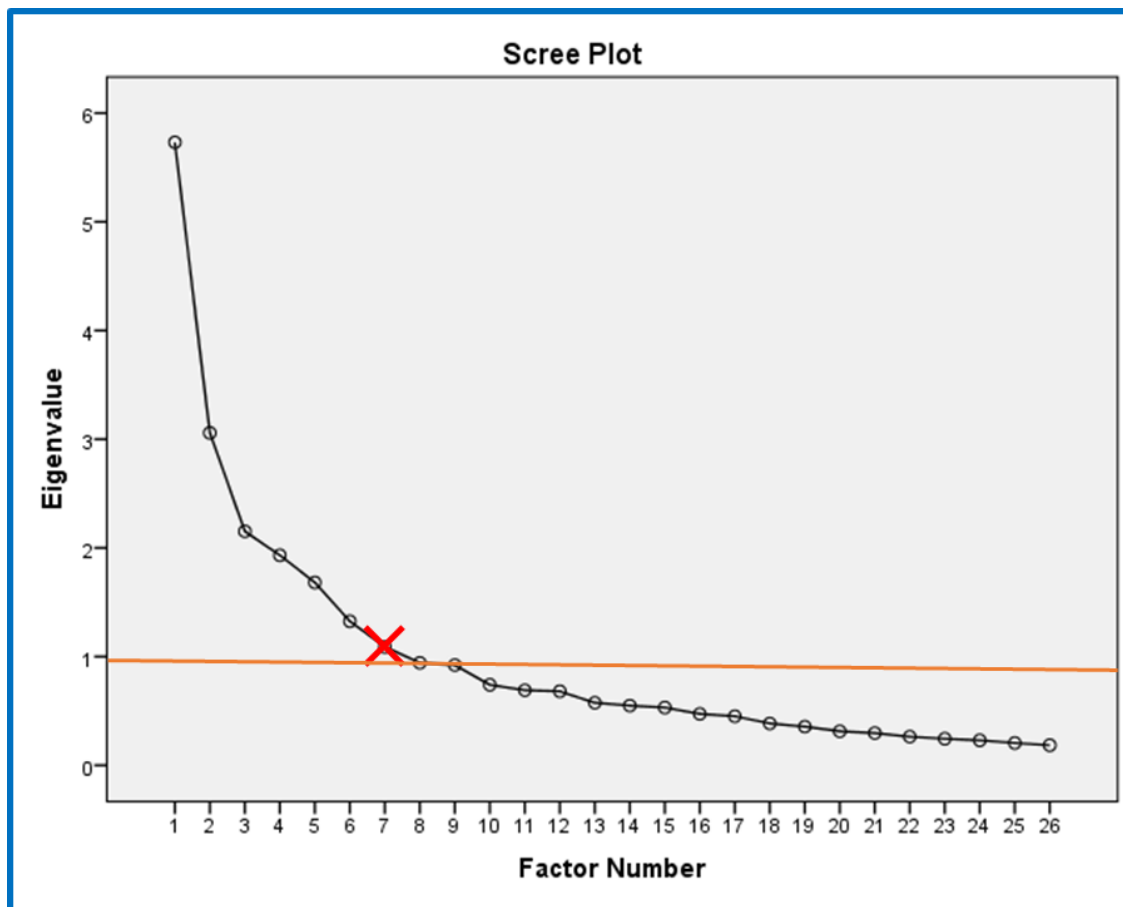
KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.۷۶۰
Bartlett's Test of Sphericity	Approx. Chi-Square	۱۸۳۳,۲۵۵
	Df	۳۲۵
	Sig.	.۰۰۰

برای انتخاب تعداد عامل‌های مناسب از جدول واریانس‌های تشریح شده و نمودار سنگ‌ریزه استفاده کردیم. معیار سرانگشتی کیسِر بیان می‌کند تنها عامل‌هایی باقی بمانند که مقدار ویژه آنها یک و بیش از یک است و همچنین در نمودار سنگ‌ریزه بارسم خط  $y=1$  نقطه که نمودار قطع می‌شود تعداد عامل‌های قبل از آن را برای انتخاب تعداد عامل مناسب انتخاب می‌کنند. با توجه به نتایج جدول و نمودار سنگ‌ریزه (اسکری تست) در می‌بایم تعداد عامل مناسب برای مجموع دادگان ما هفت عدد است.

جدول ۳-۴ واریانس‌های تشریح شده جهت انتخاب تعداد عامل مناسب

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
۱	۵,۷۳۲	۲۲,۰۴۶	۲۲,۰۴۶	۵,۲۷۷	۲۰,۲۹۵	۲۰,۲۹۵	۲,۸۵۰	۱۰,۹۶۱	۱۰,۹۶۱
۲	۳,۰۵۸	۱۱,۷۶۲	۳۳,۸۰۹	۲,۶۲۷	۱۰,۱۰۴	۳۰,۳۹۹	۲,۰۹۰	۸,۰۳۷	۱۸,۹۹۷
۳	۲,۱۵۲	۸,۲۷۶	۴۲,۰۸۵	۱,۷۲۹	۶,۶۴۹	۳۷,۰۴۹	۲,۰۴۱	۷,۸۵۰	۲۶,۸۴۷

۴	۱,۹۳۲	۷,۴۳۲	۴۹,۵۱۷	۱,۴۷۶	۵,۶۷۶	۴۲,۷۲۴	۲,۰۰۵	۷,۷۱۲	۳۴,۵۶۰
۵	۱,۶۸۲	۶,۴۶۸	۵۵,۹۸۴	۱,۲۵۷	۴,۸۳۶	۴۷,۵۶۱	۱,۷۳۳	۶,۶۶۷	۴۱,۲۲۶
۶	۱,۳۲۵	۵,۰۹۶	۶۱,۰۸۰	.۸۶۹	۳,۳۴۲	۵۰,۹۰۲	۱,۵۹۸	۶,۱۴۶	۴۷,۳۷۲
۷	۱,۰۸۹	۴,۱۸۹	۶۵,۲۶۹	.۵۹۷	۲,۲۹۶	۵۳,۱۹۸	۱,۵۱۵	۵,۸۲۶	۵۳,۱۹۸
۸	.۹۴۳	۳,۶۲۵	۶۸,۸۹۴						
۹	.۹۲۲	۳,۵۴۵	۷۲,۴۳۹						
۱۰	.۷۴۰	۲,۸۴۷	۷۵,۲۸۶						



شکل ۴-۴ نمودار سنگ ریزه برای انتخاب تعداد عامل مناسب

دو جدول بعد نتایج تفسیر عاملی و بار عاملی‌ها را برای هر کدام از متغیرها (فعالیت‌های گردشگری) نمایش می‌دهد. جدول اول قبل از چرخش واریمکس و جدول بعدی بعد از چرخش واریمکس است که از نظر ما قابلیت تفسیر پذیری بهتری دارد.

جدول ۴-۴ جدول عامل ها به همراه بار عاملی برای هر متغیر قبل از چرخش واریمکس

Factor Matrix <sup>a</sup>							
	Factor						
	۱	۲	۳	۴	۵	۶	۷
garden	.۶۵۹						
Mansion	.۶۵۶	-.۴۵۱					
palace	.۵۷۹						
mosque	.۵۵۰						
WaterfallsAndRivers	.۵۳۶						
parks	.۵۳۱						
culturalMusem	.۵۲۶						
lakes	.۵۱۷						
zoo	.۴۶۴						
traMa	.۴۴۴						
bookExib	.۴۳۸						
tomb	.۴۲۳						
mall	.۴۲۰						
artExib							
mountains							
NoneWaterSports		.۵۸۶	.۵۰۷				
musem	.۴۹۴	-.۵۳۹					
sportEvents		.۵۲۵	.۴۵۷				
Watersports		.۴۲۴					
cinema			.۴۷۷				
rural	.۴۲۷		-.۴۶۳				
theater							
cafe				.۵۵۲			
restaurant	.۴۳۲			.۵۰۳			
musicCon							
EmamZadeh					-.۶۵۸		

جدول ۴-۵ جدول عامل ها به همراه بار عاملی برای هر متغیر بعد از چرخش واریمکس

Rotated Factor Matrix <sup>a</sup>							
	Factor						
	۱	۲	۳	۴	۵	۶	۷
Mansion	.۸۳۷						
musem	.۷۲۶						
palace	.۶۹۷						
traMa	.۵۳۱						
culturalMuseum	.۴۷۹						
restaurant		.۷۵۵					
cafe		.۶۶۱					
mall		.۶۲۷					
musicCon							
WaterfallsAndRivers			.۶۹۳				
rural			.۶۸۸				
lakes			.۶۱۷				
mountains			.۵۹۶				
NoneWaterSports				.۹۰۴			
sportEvents				.۷۶۵			
Watersports				.۵۲۴			
theater					.۶۹۶		
bookExib					.۵۱۱		
artExib					.۴۹۵		
cinema					.۴۲۸		
EmamZadeh						.۸۵۶	
mosque						.۵۵۹	
tomb							
parks							.۵۸۰
garden							.۵۷۶
zoo							.۵۴۴

با توجه به جدول‌ها عامل‌ها بعد از چرخش واریمکس ما هفت گروه یا کلیشه گردشگری در استان تهران بدست آوردیم و آن‌ها را نام‌گذاری کردیم و پاسخی به یکی از سوالات پژوهش ما در رابطه با دسته‌ها گردشگری و تعداد آن‌ها بر مبنای فعالیت‌های گردشگری و تفریحی شناسایی شده در استان تهران است. نتایج این فرآیند به طور خلاصه در جدول ۴-۶ آمده است.

جدول ۴-۶ جدول نام‌گذاری بر روی عامل‌های حاصل از تحلیل عاملی

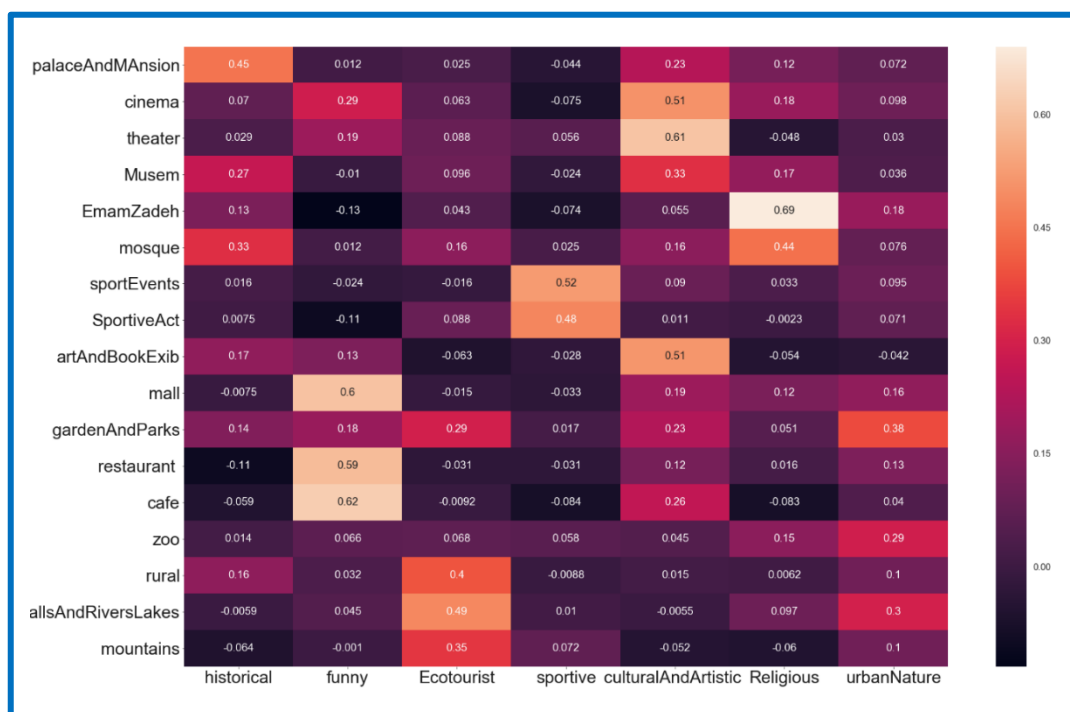
معادل انگلیسی برای تحلیل در نرم افزار	متغیرهای عامل	نام عامل
historical	عمارت گردی، موزه گردی، کاخ و قصر گردی، بازارهای سنتی و موزه‌های فرهنگ و هنر	تاریخی
funny	رستوران، کافه و بازار گردی	مفرح
Ecotourist	رودخانه و آبشارها، روستاها، دریاچه‌ها، کوهستان	طبیعت گرد
Sportive	ورزش‌های آبی، ورزش‌های غیر آبی، رویدادهای ورزشی	ورزشی
Cultural	نمایشگاه کتاب، نمایشگاه هنری، سینما، تئاتر	فرهنگی و هنر
Religious	مسجد، امام زاده‌ها	مذهبی
UrbanNature	پارک و بوستان‌ها، باغ‌های عمومی، باغ وحش	طبیعت گردی شهری

در این تحلیل، فعالیت زیارت از قبور مشاهیر و کنسرت موسیقی به دلیل کم بودن ضریب عاملی کمتر از ۰,۴ حذف شد.

#### ۴,۵ ساخت آنتولوژی

حال که هفت گروه بر اساس تحلیل عامل بدست آمده است، از این هفت گروه برای سطح دوم آنتولوژی استفاده خواهیم کرد. همچنین در سطح مفهوم در آنتولوژی که فعالیت‌های ما (برچسب نمایه کاربر) قرار دارد، ما از فعالیت‌های قبلی که برای تحلیل عاملی استفاده کردیم، با کمی تغییر برای لایه مفهوم استفاده خواهیم کرد. منظور از تغییر ادغام برخی از این فعالیت‌ها که دارای همبستگی قابل قبول با یکدیگر هستند، برای افزایش دقت الگوریتم‌های پیش‌بینی کننده است. در مجموع ۱۷ فعالیت در سطح مفهوم قرار گرفتند و برای هر کدام از آن‌ها نمونه‌هایی به عنوان مثال در سطح نمونه آورده شده است. شکل ۵-۴ بخشی از آنتولوژی ارائه شده را نمایش می‌دهد.

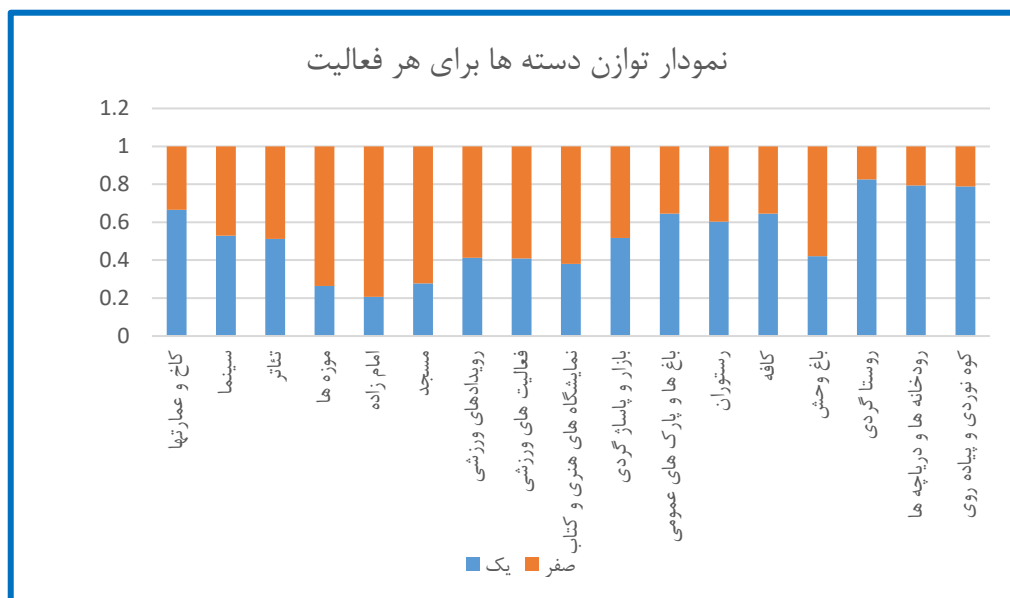




شکل ۴-۶ نمودار گرمایی ضریب همبستگی پیرسون بین گروه ها و فعالیت ها گردشگری

هدف از الگوریتم‌های دسته‌بند در این مرحله، آموزش مدلی است که بتواند بین سطح دوم آنتولوژی یعنی گروه‌های ایجاد شده از تحلیل عاملی و سطح سوم آنتولوژی موسوم به سطح مفهوم که در آن فعالیت‌های گردشگری قرار دارند را نگاشت ایجاد کند. از رویکرد تبدیل مسئله برای هر سه روش باینری رلونس، لیبیل پاورست و دسته‌بند زنجیری استفاده شده است. روش نام‌گذاری مدل‌ها در جدول ارزیابی برای باینری رلونس به صورت Br\_ModelName، برای دسته‌بند زنجیره‌ای به صورت CC\_ModelName و برای روش لیبیل پاورست به صورت LP\_ModelName است. نمودار زیر شکل توازن دسته‌ها (صفر نشان این است که فرد آن فعالیت را نخواسته است و یک نشان این است که فرد به آن فعالیت علاقه دارد). برای هر کدام از فعالیت‌ها (برچسب‌ها) نمایش داده شده است. برای رفع مشکل داده‌ها نامتوازن، برای الگوریتم کت‌بوست از روش SMOTE استفاده شده است و برای الگوریتم‌های جنگل تصادفی، درخت تصمیم، رگرسیون لجستیک، درختان اضافی و بردارهای ماشین پشتیبان از ویژگی‌ها و امکانات خود این الگوریتم‌ها برای پرداختن به داده‌ها متوازن در هر سه رویکرد تبدیل مسئله و الگوریتم‌های یادگیرنده گروهی استفاده شده است. الگوریتم‌های تطبیق‌پذیر ضعف امکان مدیریت داده‌های متوازن را ندارند و الگوریتم‌های کا-نزدیکترین همسایه و بیز ساده مشکلی در پرداختن به این دسته از داده‌ها ندارند. از طرفی همانطور که گفتیم داده‌های نامتوازن ممکن است نتایج گمراه‌کننده داشته باشد به همین دلیل از

پنج معیار دقت، صحت، بازآوری، امتیاز اف و در آخر تابع ضرر همینگ استفاده کرده‌ایم که تا نتایج قابل اتکا بدست آوریم.



شکل ۴-۷ نمودار شکل توازن دسته ها برای هر کدام از فعالیت ها

## ۴.۷ ارزیابی برون خطی

هدف از این ارزیابی، بررسی عملکرد الگوریتم‌های دسته‌بند در ایجاد نگاشت بین گروه‌های گرد شگری و فعالیت‌های گرد شگری است. الگوریتم‌ها در این بخش مورد ارزیابی و مقایسه با هم قرار می‌گیرند. نتایج این ارزیابی‌ها در جدول ۴-۷ و نمودار ۴-۸ نمایش داده شده و قیاسی از عملکرد این الگوریتم‌ها در پیش‌بینی فعالیت‌ها برای کاربر می‌دهند. نکته‌ای که در این قسمت باید آن اشاره کنیم میزان کم بودن تابع ضرر همینگ نشان عملکرد خوب الگوریتم و در سایر معیارهای ارزیابی، زیاد بودن این معیارها نشان عملکرد بالای الگوریتم است. همچنین بازه امتیازات بین صفر و یک است. تمامی محاسبات توسط بسته‌های نرم افزاری زبان برنامه نویسی پایتون موسوم به سایکیت‌لرن<sup>۳</sup> و سایکیت‌مولتی‌لرن<sup>۴</sup> که دومی بسته الحاقی به بسته اول جهت انجام دسته‌بندی‌های چندبرچسب است، انجام گرفته شده است.

<sup>۳</sup> scikit-learn

<sup>۴</sup> scikit-multilearn

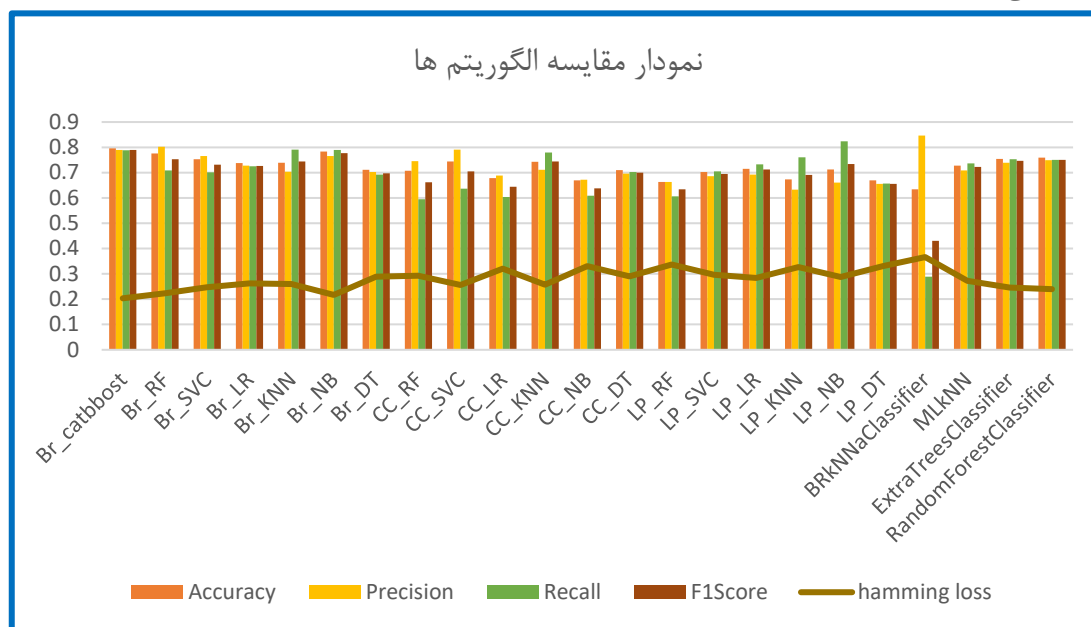


جدول ۷-۲- مقایسه عددی مدل های دسته بندی جهت پیشبینی فعالیت مورد نظر افراد

تابع ضرر همینگ	امتیاز اف	بازآوری	صحت	دقت	الگوریتم	رویکرد
۰٫۲۰۳۴۷۱۵۵	۰٫۷۸۹۴۲	۰٫۷۸۸۷۳	۰٫۷۹۰۱	۰٫۷۹۶۵۲۸	Br_catbboost	باینری رلونس
۰٫۲۲۳۷۲۲۲۸	۰٫۷۵۲۶۷	۰٫۷۰۸۸۴	۰٫۸۰۲۲۷	۰٫۷۷۶۲۷۸	Br_RF	
۰٫۲۴۶۸۶۵۹۶	۰٫۷۳۱۶۶	۰٫۷۰۰۸	۰٫۷۶۵۳۵	۰٫۷۵۳۱۳۴	Br_SVC	
۰٫۲۶۲۲۹۵۰۸	۰٫۷۲۶۳۶	۰٫۷۲۴۹	۰٫۷۲۷۸۲	۰٫۷۳۷۷۰۵	Br_LR	
۰٫۲۶۰۳۶۶۴۴	۰٫۷۴۴۴۸	۰٫۷۹۱۱۶	۰٫۷۰۳۵۷	۰٫۷۳۹۶۳۴	Br_KNN	
۰٫۲۱۶۹۷۲۰۳	۰٫۷۷۷۴۵	۰٫۷۸۹۱۶	۰٫۷۶۶۰۸	۰٫۷۸۳۰۲۸	Br_NB	
۰٫۲۸۸۳۳۱۷۳	۰٫۶۹۷۶۷	۰٫۶۹۲۷۷	۰٫۷۰۲۶۵	۰٫۷۱۱۶۶۸	Br_DT	
۰٫۲۹۲۱۸۹۰۱	۰٫۶۶۱۴۵	۰٫۵۹۴۳۸	۰٫۷۴۵۵۹	۰٫۷۰۷۸۱۱	CC_RF	دسته بند زنجیره ای
۰٫۲۵۵۵۴۴۸۴	۰٫۷۰۵۲۳	۰٫۶۳۶۵۵	۰٫۷۹۰۵۲	۰٫۷۴۴۴۵۵	CC_SVC	
۰٫۳۲۱۱۱۸۶۱	۰٫۶۴۳۸۵	۰٫۶۰۴۴۲	۰٫۶۸۸۷۹	۰٫۶۷۸۸۸۱	CC_LR	
۰٫۲۵۷۴۷۳۴۸	۰٫۷۴۴۰۱	۰٫۷۷۹۱۲	۰٫۷۱۱۹۳	۰٫۷۴۲۵۲۷	CC_KNN	
۰٫۳۳۰۷۶۱۸۱	۰٫۶۳۸۵۷	۰٫۶۰۸۴۳	۰٫۶۷۱۸۴	۰٫۶۶۹۲۳۸	CC_NB	
۰٫۲۹۰۲۶۰۳۷	۰٫۶۹۹۳	۰٫۷۰۲۸۱	۰٫۶۹۵۸۳	۰٫۷۰۹۷۴	CC_DT	
۰٫۳۳۶۵۴۷۷۳	۰٫۶۳۳۷۹	۰٫۶۰۶۴۳	۰٫۶۶۳۷۴	۰٫۶۶۳۴۵۲	LP_RF	لیبل پاورست
۰٫۲۹۷۰۱۰۶۱	۰٫۶۹۵۰۵	۰٫۷۰۴۸۲	۰٫۶۸۵۵۵	۰٫۷۰۲۹۸۹	LP_SVC	
۰٫۲۸۴۴۷۴۴۵	۰٫۷۱۲۲	۰٫۷۳۲۹۳	۰٫۶۹۲۶	۰٫۷۱۵۵۲۶	LP_LR	
۰٫۳۲۶۹۰۴۵۳	۰٫۶۹۰۹۸	۰٫۷۶۱۰۴	۰٫۶۳۲۷۲	۰٫۶۷۳۰۹۵	LP_KNN	
۰٫۲۸۷۳۶۷۴۱	۰٫۷۳۳۴۵	۰٫۸۲۳۲۹	۰٫۶۶۱۲۹	۰٫۷۱۲۶۳۳	LP_NB	
۰٫۳۳۰۷۶۱۸۱	۰٫۶۵۵۹۷	۰٫۶۵۶۶۳	۰٫۶۵۵۳۱	۰٫۶۶۹۲۳۸	LP_DT	
۰٫۳۶۶۴۴۱۶۶	۰٫۴۳۱۱۴	۰٫۲۸۹۱۶	۰٫۸۴۷۰۶	۰٫۶۳۳۵۵۸	BRkNNaClassifier	یادگیری تطبیق- پذیر
۰٫۲۷۱۹۳۸۲۸	۰٫۷۲۲۴۴	۰٫۷۳۶۹۵	۰٫۷۰۸۴۹	۰٫۷۲۸۰۶۲	MLkNN	
۰٫۲۴۵۵۹۰۱۶۴	۰٫۷۴۶۲۷	۰٫۷۵۳۰۱	۰٫۷۳۹۶۴	۰٫۷۵۴۰۹۸	ExtraTreesClassifier	یادگیری گروهی
۰٫۲۴۰۱۱۵۷۲	۰٫۷۵۰۲۵	۰٫۷۵۱	۰٫۷۴۹۵	۰٫۷۵۹۸۸۴	RandomForestClassifier	

با نگاهی به جدول بالا الگوریتم کت بوست با رویکرد باینری رلونس بالاترین دقت را داشته است و همچنین میزان صحت و بازآوری آن نیز نزدیک به دقت آن است که دلیل آن بیش نمونه گیری مناسب داده ها است و همچنین بالاترین امتیاز اف را در بین الگوریتم ها دارا است. اما با این وجود این برتری جزئی است چرا که الگوریتم های جنگل تصادفی و بیز ساده نیز عملکرد خوبی از خودشان نشان داده اند. در رویکرد دسته بند زنجیره ای عملکرد الگوریتم ها و دقت آن ها افت بسیار محسوسی داشته است و دلیل آن نیز این است که با توجه به نمودار همبستگی که ارائه کردیم، تمام فعلیت های با یکدیگر همبستگی بالا نداشتند و فقط صرفا برخی از آن ها ارتباط و همبستگی مناسب داشتند. به همین دلیل در رویکرد زنجیره ای اضافه شدن

متغیرها به صورت زنجیری به مدل دسته‌بند صرفاً موجب افزایش ابعاد مسئله و کاهش دقت مدل دسته‌بند می‌شود.



شکل ۴-۸ نمودار مقایسه عملکرد دسته‌بندها جهت پیش‌بینی فعالیت مورد نظر کاربران

در رویکرد لیبل‌پاورست نیز عملکرد مانند دسته‌بندهای زنجیره‌ای است. علت ضعف عملکرد در این رویکرد می‌تواند به این دلیل باشد که تعداد فعالیت‌ها که الگوریتم باید پیش‌بینی کند زیاد است و زمانی که در رویکرد لیبل‌پاورست مسئله را تبدیل به یک مسئله دسته‌بند یک برچسب می‌کند تعداد دسته‌های موجود در برچسب زیاد است و این در یادگیری الگوریتم مشکل ایجاد می‌کند. در رویکرد الگوریتم‌های تطبیق‌پذیر الگوریتم BRkNNaClassifier در مواجهه با داده‌های نامتوازن عملکرد بسیار بد و ناامید کننده‌ای داشت ولی الگوریتم MLkNN عملکرد نسبتاً بهتری داشته ولی در قیاس با سایر رویکردها و به خصوص باینری رلونس بسیار ضعیف‌تر عمل کرده است. اما آخرین رویکرد الگوریتم‌های یادگیرنده گروهی، هر دو الگوریتم جنگل تصادفی و درختان زیادی عملکرد نسبتاً مشابهی داشتند و بعد از رویکرد باینری رلونس بهترین نتایج را ارائه دادند و شاید بتوان اینطور استنباط کرد که الگوریتم‌هایی که ارتباط بین فعالیت‌های گردشگری را در نظر نگرفته و با آنها به صورت یک مسئله جدا برخورد می‌کنند عملکرد بهتری دارند. از جهت مقایسه باید اشاره کنیم در رویکردهای دسته‌بند چند برچسب مقاله (Rivoli et al., ۲۰۱۷) با پرسش ۲۱ سوال و پیش‌بینی ۱۲ نوع غذا برای کاربر، در بهترین حالت برای امتیاز اف کوچک مقدار ۰,۵۴۰ را در رویکرد باینری رلونس بدست آورده و مقاله (Zheng et al., ۲۰۱۴) برای

دادگان تریپادوایزر، با دریافت شش سوال از کاربر و پیش‌بینی ۵ نوع کاربری، در بهترین حالت در دقت به امتیاز نزدیک به ۰,۸۵ و در صحت نیز نزدیک به ۰,۸۵ است. از جهت برتری رویکرد ما نسبت به مقاله اول، تعداد پاسخ سوالات کاربر به پرسش‌نامه و دریافت اطلاعات هفت سوال از او است و همچنین عملکرد کارتری در ارزیابی‌ها داشته است و برتری آن نسبت به مقاله دوم، تعداد پیش‌بینی‌های متنوع (۱۷ پیش‌بینی) نسبت به آن دارد و این موجب می‌شود فرآیند شخصی سازی اطلاعات بسیار شود. از جهت معیارهای عددی نیز اختلاف ما با مقاله مورد اشاره نزدیک به هفت درصد است که با توجه به طیف متنوع پیش‌بینی‌ها و شخصی سازی بیشتر اطلاعات برای کاربر کاملاً توجیه‌پذیر است. همچنین با نگاه عملکرد الگوریتم ک-نزدیکترین هم‌سایه که در سامانه‌های پیشنهادگر پراستفاده هستند، در مقایسه با سایر الگوریتم‌ها عملکرد نسبتاً متوسط را ارائه می‌دهد و در روش‌های تطبیق‌پذیر به طور مثال الگوریتم BRkNNaClassifier عملکرد ناامید کننده‌ای در امتیاز بازآوری دارد. همانطور که پیش‌تر گفتیم عدم توانایی در برخورد با داده‌های نامتوازن و همچنین تنک بودن داده‌ها می‌تواند موجب عملکرد ضعیف این الگوریتم باشد.

#### ۴,۸ تشریح عملکرد دسته‌بند

حال قصد پاسخ به سوال دیگری از پژوهش در رابطه با میزان اثر گذاری امتیاز دهی دسته‌ها توسط کاربر بر انتخاب شدن فعالیت برای او است. در این بخش الگوریتم کت‌بوست به دلیل عملکرد بهتر نسبت به سایر الگوریتم‌ها جهت بررسی انتخاب شده است. جدول ۴-۸ زیر اهمیت هر کدام از گروه‌ها را در پیش‌بینی فعالیت‌ها نمایش می‌دهد. فاصله اعداد بین صفر و یک هستند و جمع ضرایب اهمیت برای یک فعالیت برابر یک است. هر چه ضریب اهمیت به یک نزدیک‌تر باشد، آن گروه برای پیش‌بینی فعالیت مورد نظر بیشترین تاثیرگذاری را دارد. به طور مثال برای این که به شخص توسط سامانه موزه‌گردی پیشنهاد شود، امتیاز آن شخص به گروه‌های تاریخ و فرهنگ‌و هنر به ترتیب بیشترین تاثیر را دارند.

جدول ۴-۸ نمایش عددی گروه‌های تاثیر گزار در پیش‌بینی فعالیت‌ها توسط مدل‌های دسته‌بند کت‌بوست

طبیعت گرد شهری	مذهبی	فرهنگی و هنری	ورزشی	طبیعت گرد	مفرح	تاریخی
۰,۰۰۸۶	۰,۰۶۸۸	۰,۰۲۲۲	۰,۱۱۶۳	۰,۰۶۹۶	۰,۱۰۵۰	۰,۶۰۹۶
۰,۱۲۴۸	۰,۱۱۰۴	۰,۳۹۸۰	۰,۱۱۷۴	۰,۰۲۷۲	۰,۱۰۸۸	۰,۱۱۳۳
۰,۱۲۵۸	۰,۰۹۸۴	۰,۴۱۹۷	۰,۰۷۴۸	۰,۰۴۷۰	۰,۱۱۸۳	۰,۱۱۵۹
۰,۱۰۵۵	۰,۱۹۳۲	۰,۲۲۳۴	۰,۰۹۱۲	۰,۰۹۱۹	۰,۰۶۱۹	۰,۲۳۲۸
۰,۱۰۵۱	۰,۵۱۵۶	۰,۰۹۱۲	۰,۰۷۳۱	۰,۰۳۷۳	۰,۱۰۷۹	۰,۰۶۹۸
۰,۱۳۱۴	۰,۲۱۸۱	۰,۱۵۲۰	۰,۱۲۳۷	۰,۰۴۷۵	۰,۱۴۴۹	۰,۱۸۲۴
۰,۱۳۹۶	۰,۰۶۰۴	۰,۰۹۱۵	۰,۴۱۰۳	۰,۰۳۵۴	۰,۱۴۳۰	۰,۱۱۹۹

فعالیت های ورزشی	۰,۱۸۰۶	۰,۰۲۱۲	۰,۰۵۹۱	۰,۰۵۱۵۸	۰,۰۷۷۵	۰,۰۴۳۲	۰,۱۰۲۶
نمایشگاه های هنری و کتاب	۰,۱۴۷۹	۰,۰۸۶۸	۰,۱۰۵۶	۰,۱۰۴۳	۰,۲۹۰۳	۰,۱۴۹۳	۰,۱۱۵۸
بازار گردی	۰,۰۷۸۷	۰,۴۳۹۳	۰,۰۴۸۹	۰,۱۰۰۹	۰,۱۰۹۲	۰,۱۳۴۹	۰,۰۸۸۱
پارک ها و باغ های عمومی	۰,۱۳۶۵	۰,۱۵۷۷	۰,۱۲۱۸	۰,۱۷۶۶	۰,۱۱۹۹	۰,۱۱۶۹	۰,۱۷۰۶
رستوران	۰,۰۷۵۰	۰,۴۹۰۹	۰,۰۱۹۴	۰,۰۵۵۵	۰,۰۶۷۳	۰,۱۴۵۵	۰,۱۴۶۴
کافه	۰,۱۴۲۴	۰,۴۰۴۸	۰,۰۵۹۳	۰,۰۶۵۳	۰,۱۱۰۷	۰,۱۴۳۸	۰,۰۷۳۶
باغ وحش	۰,۰۶۷۸	۰,۱۳۵۱	۰,۰۵۵۰	۰,۱۷۵۲	۰,۱۱۷۳	۰,۲۱۷۱	۰,۲۳۲۵
روستا گردی	۰,۱۱۷۲	۰,۱۰۴۸	۰,۲۹۴۱	۰,۱۰۳۱	۰,۱۰۴۹	۰,۱۳۸۸	۰,۱۳۷۱
رودخانه و دریاچه ها	۰,۰۳۸۱	۰,۰۲۸۸	۰,۴۶۰۷	۰,۱۲۵۱	۰,۰۹۸۱	۰,۰۹۶۳	۰,۱۵۳۰
کوهنوردی	۰,۱۰۸۹	۰,۰۵۷۸	۰,۳۴۲۶	۰,۰۹۶۷	۰,۰۷۳۲	۰,۰۴۱۳	۰,۲۷۹۵

#### ۴,۹ ارزیابی به صورت مبتنی بر سناریو

بر طبق بررسی های پرسشنامه دوم، به طور متوسط هر کاربر ۹-۸ فعالیت از ۱۷ فعالیت را انتخاب کرده است. حال فرض کنید کاربری قصد دریافت پیشنهاد از سوی این سامانه را داشته باشد. برای این منظور باید به هر کدام هفت گروه سطح دوم آنتولوژی که با تحلیل عاملی بدست آمده است از یک تا پنج امتیاز بدهد. کابر مورد نظر امتیازات را طبق جدول زیر به هر کدام از هفت گروه داده است. به عبارتی کاربر تنها اطلاعاتی که در اختیار سامانه قرار می دهد امتیاز یک تا پنج او به این هفت گروه است.

نام گروه	امتیاز
تاریخی	۴
مفرح	۵
طبیعت گرد	۳
ورزشی	۱
فرهنگی و هنر	۱
مذهبی	۳
طبیعت گردی شهری	۱

حال نتایج این کاربر وارد الگوریتم دسته بند چند برچسب که با مجموعه ای از داده های قبلی توسط پرسشنامه دوم آموزش دیده شده است، می شود. در این مرحله الگوریتم فعالیت ها (برچسب) مورد نظر کاربر را بر اساس همان هفت گروه امتیاز داده شده پیش بینی می کند. فعالیت ها پیش بینی شده برای کاربر توسط الگوریتم بازدید از کاخ ها و عمارت ها، پاساژ و بازارگردی، باغ ها و پارک های عمومی، کافه

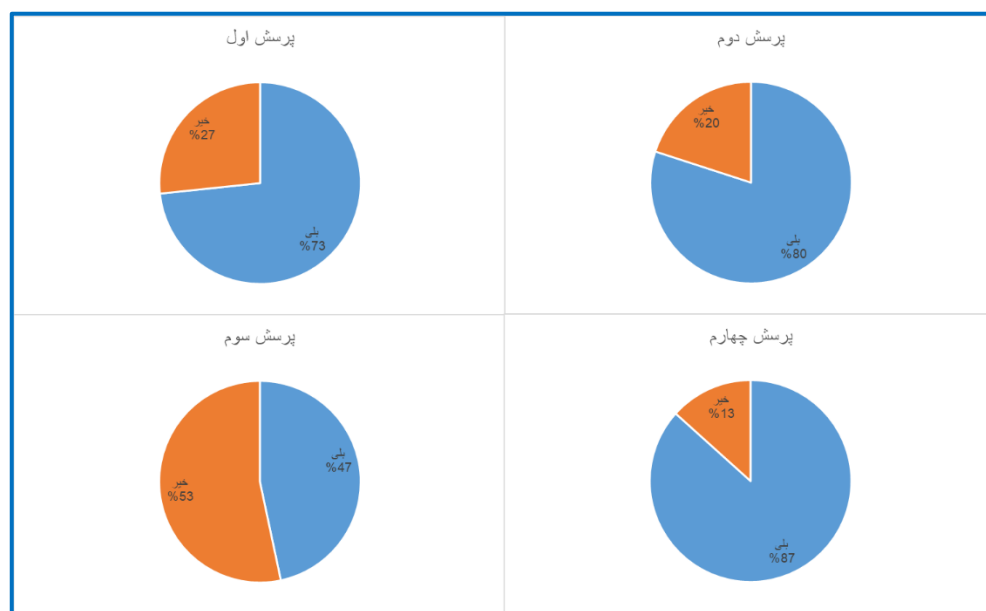
رفتن، کوهنوردی و گشت و گذار در کنار دریاچه‌ها و رودخانه‌ها و روستاها است. هر کدام از این فعالیت‌ها در سطح آخر آنتولوژی دارای نمونه‌هایی هستند که کاربر می‌تواند آن‌ها را برای گردشگری و تفریح انتخاب کند.

#### ۴.۱۰ ارزیابی مطالعه کاربر

برای ارزیابی مطالعه کاربر از روش آزمایشگاهی/کارگاهی استفاده کردیم و در سه نوبت و هر نوبت پنج نفر را مورد ارزیابی قرار دادیم. ابتدا از افراد خواستیم که به هفت گروه گردشگری امتیاز بدهند، سپس این امتیازات را به مدل آموزش داده شده دادیم و مدل فعالیت‌های مورد علاقه کاربران را پیشبینی کرد، سپس بر اساس خروجی مدل و بانک داده‌ای که از فعالیت‌های گردشگری در قالب یک آنتولوژی جمع‌آوری کرده بودیم ماشین به تصادف از هر فعالیت دو نمونه عینی در یک رابط کاربری ساده و توضیح کوتاه از هر مکان، به کاربر پیشنهاد کرد و کاربر با کلیک بر روی هر پیشنهاد می‌توانست اطلاعات کامل‌تری از فعالیت مورد نظر بدست آورد. بعد از مشاهده نتایج توسط کاربر، از او خواستیم به پرسش‌های زیر در یک فرم پرسشنامه کاغذی پاسخ دهد. پاسخ‌های ممکن برای هر پرسش بلی یا خیر بود.

۱. آیا پیشنهادات ارائه شده مطابق با نظر کاربر است؟
۲. آیا پیشنهادات ارائه شده انگیزه جهت انجام فعالیت‌های مورد پیشنهاد را افزایش داده است؟
۳. آیا افراد مایل هستند از برنامه‌ای که مبتنی بر پیشنهادگر مکان و محتوای گردشگری است برای برنامه‌ریزی گردش و تفریح خود استفاده کنند؟
۴. آیا برچسب‌های پیشنهادی ارائه شده در این سامانه از تنوع کافی برخوردار است؟

نتایج پاسخ به پرسش‌های و ارزیابی پیشنهادگرها توسط افراد به صورت زیر است:



شکل ۹-۴ نتایج پاسخ به پرسش‌ها در ارزیابی مطالعات کاربر

همانطور که از نتایج ارزیابی مشاهده می‌کنیم، میزان رضایت از پیشنهادات مناسب بوده، همچنین می‌توان استنباط کرد که این سامانه موجب انگیزش در افراد جهت انجام فعالیت‌های گردشگری می‌شود. عملکرد سامانه در ارائه پیشنهادات متنوع، همانطور که از اهداف این پژوهش نیز است، در حد مطلوبی ارزیابی شده اما در طرف دیگر میزان اعتماد افراد به پیشنهادگر مورد بحث، ضعیف است و کمتر از نصف افراد حاضر در آزمایش، علاقه به برنامه‌ریزی گردشگری با سامانه پیشنهادگر را دارند. در قسمت پرسش‌های پژوهش مطرح شده بود که آیا نتایج مطالعه کاربری و برون خطی نتایج مشابهی بدست می‌آورند؟ و با توجه نتایج مطالعه کاربری و برون خطی با در نظر گرفتن تعداد کم بودن نمونه‌های آزمایش و بایاس بودن نتیجه‌ها، می‌توان گفت در پژوهش ما نتایج هر دو ارزیابی تقریباً مشابه می‌باشد.

## ۴,۱۱ خلاصه فصل

در این فصل ابتدا ۲۶ فعالیت گردشگری در تهران را شناسایی کردیم، سپس به جمع‌آوری پرسشنامه جهت شناسایی علاقه کاربران نسبت به هر کدام از این ۲۶ فعالیت پرداختیم. نتایج پرسشنامه را با تحلیل عاملی تحلیل کردیم و در مجموع هفت گروه گردشگری تعریف کردیم. در مرحله بعد تعدادی از فعالیت‌ها که دارای همبستگی بالایی بودند را با یکدیگر ادغام کردیم و در مجموع ۱۷ فعالیت گردشگری باقی ماند و سپس به طراحی بخشی از آنتولوژی پرداختیم و مجدداً جهت کشف ارتباط علاقه افراد به گروه‌ها گردشگری و فعالیت‌ها اقدام به تهیه پرسشنامه کردیم. نتایج پرسشنامه را برای آموزش ۲۳ مدل دسته‌بند چند برچسب استفاده کردیم که رویکرد باینری رلونس بهترین امتیازات در عملکرد داشتند. بعد با

ارائه جدولی میزان تاثیرگذاری هر کدام از گروه‌های گرد شگری در انتخاب شدن یک فعالیت را از جهت اهمیت بررسی کردیم. سپس یک کاربر فرضی را در نظر گرفتیم و با رویکردی شبیه سازی/ سناریوگونه امتیازات یک کاربر فرضی به هر کدام از گروه‌ها را به مدل دادیم و نتایج و فعالیت‌های پیش‌بینی شده سامانه به کاربر را گزارش کردیم. در مرحله ارزیابی عملی، از مطالعه کاربری از روش آزمایشگاهی استفاده کردیم و سه گروه پنج نفره از افراد را مورد بررسی قرار دادیم و در آخر خواستیم به سوالات در یک پرسشنامه کاغذی پاسخ دهند که نتایج این پرسشنامه را نیز در همین فصل گزارش کردیم.

## **فصل پنجم**

### **جمع‌بندی و نتیجه‌گیری**



## ۵,۱ مقدمه

هدف از این فصل، مروری بر فصل‌های گذشته و ارائه خلاصه‌ای از فعالیت‌های انجام شده است. سپس استنباط و نتیجه‌گیری خود را از این پژوهش و دستاوردهای آن را شرح خواهیم داد و در آخر نگاهی به محدودیت‌های پژوهش حاضر و پیشنهادات برای پژوهش‌های آتی خواهیم داشت.

## ۵,۲ مرور و تحلیلی بر فصل‌های گذشته

در این پژوهش تشریح کردیم که انسان امروزی با توجه به امکانات و توسعه‌های فناوری جابه‌جایی و تحرک بیشتری دارد. گردشگری یکی از بزرگترین و پر درآمدترین صنایع دنیاست، اما متأسفانه کشورمان ایران با توجه به توانایی‌ها و ویژگی‌های بالقوه‌ای که دارد آن‌چنان که باید از این فرصت بهره نگرفته است. در این میان فناوری اطلاعات نقش بسیار موثر و چشمگیری در توسعه گردشگری دارد. اینترنت سبب مزیت دو سویه بین کاربران و ارائه دهندگان خدمات شده است. اما با تمام این مزیت‌ها، مقادیر انبوه اطلاعات در سراسر وب خود موجب سردرگمی کاربران شده است که در این میان سامانه‌های پیشنهادگر روشی برای پرداختن به این مشکل است. پیشنهادگرها به طور کلی دارای چهار رویکرد کلی پالایشگر مبتنی بر محتوا، پالایشگر مبتنی بر دانش، پالایشگر مشارکتی و نوع ترکیبی از موارد ذکر شده است. پیشنهادگرها از روش‌ها و فن‌های متنوعی نیز استفاده می‌کنند. روش‌های یادگیری ماشین، کلیشه‌ها، بافت‌آگاهی از متن و ... بخشی از این فن‌ها است. در این میان یکی از مهم‌ترین این فن‌ها که ما در این پژوهش از آن بهره بردیم، آنتولوژی است که برای تشریح یک دامنه مشترک به طو مفهومی و روشن است. به طور کلی برای ارزیابی پیشنهادگرها از سه روش عمده ارزیابی برون خطی، ارزیابی بر خط و مطالعه کاربران استفاده می‌کنند. البته بسیاری از مقالات از نوع دیگری موسوم به شبیه‌سازی یا سناریو که فرآیندها وارد شدن کاربر به سامانه تا دریافت پیشنهادات را تشریح می‌کند، استفاده می‌کنند. پیشنهادگرها البته دارای مشکلاتی هستند و چالش‌هایی در طراحی آن‌ها وجود دارد. مشکلاتی مانند

شروع سرد، تنک بودن داده‌ها، تکرارگرایی، گوسفند خاکستری، اثر پورتفلیو، مشکل در دریافت اطلاعات دموگرافیک و پیشنهادات خیلی کم یا پیشنهادات خیلی زیاد از این دست مشکلات و چالش‌های سامانه‌های پیشنهادگر هستند. پیشنهادگرها در گردشگری به دو صورت پیشنهاد مقصد و یک برنامه سفری و نوع پیشنهاد مکان‌های گردشگری و تفریحی در یک مکان خاص است که نوع دوم در ادبیات پیشنهادگرهای گردشگری طرفدار بیشتری دارد و پژوهشگران بدان بیشتر پرداخته‌اند. کلیشه‌ها و به بیانی دیگر نقش‌ها در پیشنهادگرهای گردشگری نقش مهمی داشتند و پژوهش‌های مناسبی پیرامون آن‌ها صورت گرفته است. آنچه که نتایج پژوهش‌ها پیداست می‌توان فعالیت‌های گردشگری را بر اساس نقش‌ها و یا کلیشه‌ها که به افراد تعلق می‌گیرد برای آن‌ها پیش‌بینی کرد. در بررسی ادبیات پیشنهادگرهای گردشگری، روشن است که در ارزیابی مطالعه کاربر پراستفاده‌ترین روش برای ارزیابی سامانه پیشنهادگر گردشگری است، شاید دلیل آن استفاده از ابزارها و فن‌هایی است که معیار متریک مناسبی ارائه نمی‌کند. به طور مثال روش‌های مبتنی بر بافت آگاه و حتی در یادگیری ماشین مانند خوشه‌بندی، معیار متریک روشن و دقیقی جهت پیش‌بینی نیازهای کاربر نمی‌دهد. یکی از فن‌هایی که ما در این پژوهش برای طراحی سامانه خود استفاده کردیم یادگیری ماشین بود. یادگیری ماشین روشی برای یادگیری مدل‌ها از مجموعه‌ای از مشاهدات و سپس تعمیم آن به داده‌های جدید جهت پیش‌بینی است. یادگیری ماشین دارای دو رویکرد با نظارت و بدون نظارت است. در رویکرد با نظارت داده‌ها دارای برچسب هستند و هدف آموزش مدل برای درک و ایجاد نگاشت بین داده‌ها و برچسب‌ها است. اما در رویکرد بی‌نظارت که خود شامل دو روش خوشه‌بندی و کاهش ابعاد است، داده‌ها فاقد برچسب هستند. در روش کاهش ابعاد، تعداد متغیرهای مسئله جهت حل بهینه و افزایش دقت کاهش پیدا می‌کند اما در روش خوشه‌بندی داده‌ها بر اساس شباهت ذاتی‌شان در خوشه‌هایی با یکدیگر تجمع می‌شوند. در بخش دیگر پژوهش به تحلیل عاملی که برای خلاصه سازی داده‌ها جهت کشف الگوها و فهم بهتر داده‌ها است پرداختیم. به معیارها و روش‌های تحلیل عاملی پرداختیم و بیان کردیم که شرط این که یک تحلیل مناسب داشته باشیم چه می‌باشد. سپس در بخش مدل‌های دسته‌بند به معرفی مدل‌های، بیز، رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، درختان اضافی، بردارهای ماشین پشتیبان، کاندیدترین همسایه و الگوریتم کت بوست پرداختیم. در ادامه روش‌های ارزیابی مدل‌های دسته‌بند مانند دقت، صحت، بازآوری و ... پرداختیم. مهم‌ترین رویکردی که ما در این پژوهش به آن پرداختیم دسته‌بندهای چندبرچسب است که در این دسته‌بندها دو یا بیش از دو برچسب هدف دارند و برای حل این دست از مسائل، سه رویکرد تبدیل مسئله که خود شامل الگوریتم‌های باینری رلونس، لیبل پاورست و دسته‌بندهای زنجیره‌ای است و رویکرد دیگر استفاده از الگوریتم‌های تطبیق‌پذیر است و در رویکرد آخر استفاده از الگوریتم‌های یادگیرنده گروهی است. پس از معرفی و پرداختن به ادبیات مروری گردشگری در فصل سوم معماری سامانه خود و قدم‌ها

طراحی این معماری را تشریح کردیم. به طور کلی برای طراحی سامانه مورد نظر نیاز به شناسایی فعالیت‌های گردشگری استان تهران، جمع‌آوری نظرات افراد در رابطه با این فعالیت‌ها و ایجاد گروه‌بندی از افراد است. سپس به سنجش ارتباط بین این گروه‌ها که سطح دوم آنتولوژی است و فعالیت‌ها که در سطح مفهوم یا به عبارتی فعالیت‌های گردشگری است، به کمک پرسشنامه پرداختیم. پس از جمع‌آوری داده‌ها به کمک رویکردها روش‌های دسته‌بند چندبرچسب که در فصل دوم تشریح کردیم، نحوه آموزش مدل‌ها را برای درک ارتباطات سطح دوم و سوم آنتولوژی که تا بتواند سامانه بر مبنای هفت گروه، ترجیح کاربر به انتخاب ۱۷ فعالیت گردشگری را پیش‌بینی کند، تشریح کردیم. سپس روش‌های ارزیابی مدل‌های دسته‌بند چندبرچسب را تشریح کردیم و از دو روش شبیه‌سازی یا مبتنی بر سناریو و روش مطالعه کاربر را نیز برای ارزیابی عملکرد سامانه انتخاب کردیم. در فصل چهارم پیاده‌سازی قدم‌های پژوهش را تشریح کردیم. ۲۶ فعالیت گردشگری در استان تهران پیدا کردیم، برای پرسشنامه اول ۲۷۲ پرسشنامه جمع‌آوری کردیم، تحلیل عاملی بر روی پرسشنامه‌ها اجرا کردیم و هفت گروه گردشگری را پیدا کردیم. با توجه به این گروه‌ها یک آنتولوژی ارائه دادیم و بر مبنای آن شروع به جمع‌آوری پرسشنامه دوم کردیم، که در مجموع ۵۷۸ پرسشنامه جمع‌آوری شد. پس از ارزیابی و تحلیل‌ها از پرسشنامه ۲۳ مدل دسته‌بند چندبرچسب را بر روی داده‌ها اجرا کردیم و عملکرد آن را با پنج معیار سنجیدیم و در ادامه تحلیلی از نحوه عملکرد دسته‌بند ارائه کردیم. سپس یک سناریو در نظر گرفتیم با یک کاربر فرضی و داده‌های کاربر فرضی را به الگوریتم داده و نتایج آن را گزارش کردیم. در آخر به سراغ ارزیابی آزمایشگاهی جهت مطالعه کاربر رفتیم و سه گروه پنج‌نفره از کاربران را مورد آزمایش قرار داده و خواستیم که به سوالات پرسشنامه در رابطه با دقت عملکرد سامانه برای پیشنهادات، تنوع پیشنهادات، ایجاد انگیزه در کاربر و همچنین علاقه کاربر به استفاده از پیشنهادگرها را سنجیدیم.

### ۵.۳ نتیجه‌گیری پژوهش

ایجاد گروه‌ها گردشگری بر مبنای فعالیت‌های گردشگری در استان تهران، هدف اولیه این پژوهش بود و با توجه امتیازات بالا در آزمون‌های تحلیل عاملی جهت بسندگی و مناسب بودن تحلیل، نشان دهنده رویکرد درست ما در ایجاد گروه‌های گردشگری است. به گونه‌ای که ما این گروه‌بندی را بدون اطلاعات دموگرافیک و روانشناختی و تنها بر مبنای فعالیت‌های گردشگری ایجاد کردیم. همانطور که طبق پژوهش‌های گذشته به آن اشاره شده بود بین نقش‌ها و انتخاب افراد همبستگی و ارتباط وجود دارد و ما در این پژوهش نیز مجموعه‌ای از گروه‌های گردشگری ایجاد کردیم و بر طبق امتیازات افراد به آن گروه‌ها فعالیت‌های مورد نظر کاربر را پیش‌بینی کردیم و این خود تایید کننده پژوهش‌های گذشته است. نکته‌ای که در اینجا حائز اهمیت است، ارتباط بین این گروه‌ها و فعالیت‌ها است که برای شناخت کاربر تنها به

دریافت نظرات کاربر در رابطه با هفت گروه نیاز است یا به عبارت ساده‌تر هفت کلیک نیاز است تا کاربر پیشنهادات را دریافت کند. رویکرد متفاوت ما در این پژوهش، عدم دریافت اطلاعات دموگرافیک و خصوصی از افراد بود که معمولاً دریافت این اطلاعات موجب نگرانی کاربر و عدم اطمینان به سامانه‌ها می‌شود. خو شبخانه در این پژوهش ما نشان دادیم برای رفع مشکل شروع سرد که به طور معمول در پژوهش‌های دیگر اطلاعات دموگرافیک و صریح از کاربر دریافت می‌شود، با دریافت امتیاز کاربر به گروه‌های گردشگری نیز می‌توان به شکل مطلوب به این چالش پرداخت و دقت عملکرد الگوریتم موید این نظر است. در پژوهش‌های گذشته نتیجه‌گیری شده بود که اگر به کاربر آزادی عمل داده شود تا خود را با بیش از یک نقش توصیف کند، اون به چندین نقش امتیاز خواهد داد. در این پژوهش با توجه به نتایج مورد نظر نیز از این رویکرد استفاده کردیم و به کاربر آزادی عمل دادیم که از بین هفت گروه گردشگری به هر کدام به دلخواه از یک تا هفت نمره دهد و نتایج پژوهش ما بیانگر این مورد بود که برای انتخاب شدن هر فعالیت به طور معمول حداقل سه مورد از گروه‌های گردشگری که کاربر به آن امتیاز داده است، تاثیر گذار است. بر مبنای همین نتیجه ارتباطات بین سطوح گروه‌ها و فعالیت‌ها در آنتولوژی تحت تاثیر قرار می‌گیرد. بنابراین اینطور استنباط می‌شود که هر کدام از گروه‌ها بر روی انتخاب فعالیت‌ها، هر چند کم تاثیر گذار هستند. مجدداً دقت عملکردهای دسته‌بند و به دنبال آن جدول ارائه گزارش اهمیت هر کدام از این ویژگی‌ها در دسته‌بندی نیز تایید کننده این مهم است. همچنین سامانه‌ای که در این پژوهش طراحی شد، به طور معمول نزدیک به هشت فعالیت از ۱۷ فعالیت را برای کاربر برمی‌گزیند بنابراین در کاهش و شخصی سازی اطلاعات همانطور که خود افراد در پرسشنامه به طور میانگین نزدیک به هشت فعالیت را انتخاب می‌کردند، عمل کرده است و این نشانه مثبتی بر یادگیری مدل‌های دسته‌بند از مجموع داده‌ها است. عملکرد سامانه در معیار ارزیابی برون خطی و به عبارت دیگر در معیارهای متریک، در رویکرد دسته‌بندهای چندبرچسب، بنا به روش استفاده شده متفاوت بود. در مجموع روش باینری رلونس بیشترین امتیازات و بهترین عملکرد را در بین روش‌ها داشت. می‌توان اینطور استنباط کرد که با توجه به همبستگی کم فعالیت‌ها با یکدیگر، اگر آن‌ها مستقل از هم در نظر گرفته و مسئله را حل کرد نتیجه بهتری دریافت خواهیم کرد. در بین مدل‌ها، مدل بیز و مدل کا-نزدیکترین همسایه که در پیشنهادگرها مورد استفاده قرار می‌گیرد، در مجموع دادگان ما هم جزء برترین مدل‌ها بوده و عملکرد مناسبی داشتند گرچه با اختلاف اندکی از الگوریتم تقویتی کت بوست عقب‌تر هستند. موردی که حتماً باید به آن بپردازیم، جدال بر سر اینکه آیا نتایج ارزیابی برون خطی و مطالعه کاربر با هم اختلاف دارند و یا نتایج مشابه به بار می‌آورند. برای پاسخ به این پرسش در پژوهش ما، در بهترین حالت دقت‌های دسته بند جهت شناسایی فعالیت‌های مورد علاقه کاربر تقریباً ۷۹ درصد بود و همانطور که در فصل چهار مشاهده کردیم امتیاز کاربران به سامانه در مطالعه کاربر تقریباً ۷۳ درصد بود که در پژوهش ما نتیجه

تقریبا مشابه بود. در رابطه با ایجاد انگیزش در کاربران، ۸۰ درصد از افراد معتقد بودند که سامانه باعث انگیزش در آن‌ها می‌شود، بنابراین می‌شود اینطور استنباط کرد که طراحی و ساخت یک سامانه پیشنهادگر تجاری می‌تواند اقدام مناسبی در جهت توسعه و رونق گردشگری کشور باشد اما در طرف دیگر، در پاسخ به سوال از کاربران که آیا مایل هستند که برنامه‌ریزی‌های خود را با یک پیشنهادگر جهت تفریح و گردش انجام دهند بیش از نیمی از افراد پاسخ منفی دادند. شاید علت این تناقض عدم وجود پیشنهادگرهای تفریحی و گردشگری مناسب و نداشتن ذهنیت دقیق و روشن کاربران از پیشنهادگرها گردشگری و تفریحی باشد. در رابطه با توانایی سامانه در ارائه پیشنهادات متنوع نتایج مطالعه کاربری، نشان از توانایی این سامانه در ارائه اطلاعات متنوع می‌دهد. تقریبا ۸۷ درصد از افراد به مناسب بودن تنوع پیشنهادات امتیاز داده‌اند و دلیل آن می‌تواند شناسایی و دسته‌بندی مناسب فعالیت‌ها در سطح مفهوم آنتولوژی باشد.

#### ۵.۴ محدودیت‌ها و پیشنهادات آتی

هر پژوهشی دارای محدودیت‌هایی است و پژوهش ما نیز از این امر مستثنی نیست و دارای محدودیت‌هایی است. یکی از مهم‌ترین محدودیت‌های پژوهش حاضر عدم استفاده از فن بافت‌آگاه به خصوص مختصات مکانی گردشگر است، به گونه‌ای که در زمان ارائه پیشنهاد مواردی از هر فعالیت موقعیت مکانی کاربر را نیز در نظر بگیر که ما این مورد را برای پژوهش‌های آتی پیشنهاد می‌کنیم. محدودیت دیگر عدم امکان متوازن سازی داده‌ها در روش‌های تطبیق‌پذیر، که در پژوهش‌هایی که مدل دسته‌بندچند برچسب استفاده کرده بودند نیز مشاهده می‌شد. محدودیت بعدی عدم وجود سامانه پیشنهادگر گردشگری و تفریحی برای استان تهران جهت مقایسه بود.

مراجع

## مراجع فارسی

۱. طراحی یک سیستم توصیه گر مسافرت، آذین محمدی، دانشگاه گیلان، دانشکده فنی و مهندسی، استاد راهنما: اکبر خداپرست حقی، صفحه ۱۷-۱۴
۲. تیمورپور، بابک؛ نجفی حیدر، ۱۳۹۴، "داده کاوی با R به همراه متن کاوی و تحلیل شبکه های اجتماعی". ویرایش ۱، تهران: مرکز تحقیقات و توسعه سازمان اتکا.
۳. دکتر مهدی، ت. & دکتر علی قلی پور، س. ۱۳۸۸. عوامل موثر بر رشد صنعت گردشگری ایران. پژوهشنامه اقتصادی، سال نهم، ۱۵۷-.
۴. میثم، م.، سمیه، ه. & میترا، ا. ۱۳۹۱. بررسی جامعه شناختی توسعه پایدار گردشگری در ایران؛ موانع، چالش ها، راهکارها. علوم اجتماعی، سال ششم، ۲۵-۵۰.

## مراجع انگلیسی

۵. ADOMAVICIUS, G. & TUZHILIN, A. ۲۰۱۱. Context-aware recommender systems. *Recommender systems handbook*. Springer.
۶. AL-HASSAN, M., LU, H. & LU, J. ۲۰۱۵. A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system. *Decision Support Systems*, ۷۲, ۹۷-۱۰۹.
۷. AL, C. & LEE, H. ۱۹۹۲. A first course in factor analysis. Hillsdale, NJ: Lawrence Erlbaum Associates.
۸. AMATRIAIN, X., JAIME, A., OLIVER, N. & PUJOL, J. M. ۲۰۱۱. Data mining methods for recommender systems. *Recommender systems handbook*. Springer.
۹. AVESANI, P., MASSA, P. & TIELLA, R. ۲۰۰۵. Moleskiing. it: a trust-aware recommender system for ski mountaineering. *International Journal for Infonomics*, ۲۰, ۱۰-۱.
۱۰. BEEL, J., GIPP, B., LANGER, S. & BREITINGER, C. ۲۰۱۶. paper recommender systems: a literature survey. *International Journal on Digital Libraries*, ۱۷, ۳۳۸-۳۰۵.
۱۱. BERGER, H., DENK, M., DITTENBACH, M., MERKL, D. & PESENHOFER, A. ۲۰۰۷. Quo Vadis Homo Turisticus? Towards a picture-based tourist profiler. *Information and Communication Technologies in Tourism ۲۰۰۷*. Springer.

١٢. BORRÀS, J., MORENO, A. & VALLS, A. ٢٠١٤. Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, ٤١, ٧٣٨٩-٧٣٧٠.
١٣. BURKE, R. & RAMEZANI, M. ٢٠١١. Matching recommendation technologies and domains. *Recommender systems handbook*. Springer.
١٤. BURKE, R. D., HAMMOND, K. J. & YOUNG, B. C. Knowledge-based navigation of complex information spaces. Proceedings of the national conference on artificial intelligence, ١٩٩٦. ٤٦٨
١٥. CIURANA SIMÓ, E. R. ٢٠١٢. *Development of a Tourism recommender system*. Universitat Politècnica de Catalunya.
١٦. CONSOLE, L., TORRE, I., LOMBARDI, I., GIORIA, S. & SURANO, V. ٢٠٠٣. Personalized and adaptive services on board a car: an application for tourist information. *Journal of Intelligent Information Systems*, ٢١, ٢٨٤-٢٤٩
١٧. COSTELLO, A. B. & OSBORNE, J. W. ٢٠٠٥. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research & evaluation*, ١٠, ٩-١
١٨. CREMONESI, P., GARZOTTO, F. & TURRIN, R. ٢٠١٢. Investigating the persuasion potential of recommender systems from a quality perspective: An empirical study. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, ٢, ١١
١٩. DOROGUSH, A. V., ERSHOV, V. & GULIN, A. CatBoost: gradient boosting with categorical features support.
٢٠. EKSTRAND, M. D., KANNAN, P., STEMPER, J. A., BUTLER, J. T., KONSTAN, J. A. & RIEDL, J. T. Automatically building research reading lists. Proceedings of the fourth ACM conference on Recommender systems, ٢٠١٠. ACM, ١٦٦-١٥٩
٢١. FESENMAIER, D. R., RICCI, F., SCHAUMLECHNER, E., WÖBER, K. & ZANELLA, C. ٢٠٠٣. *DIETORECS: Travel advisory for multiple decision styles*, na.
٢٢. GARCÍA-CRESPO, A., CHAMIZO, J., RIVERA, I., MENCKE, M., COLOMO-PALACIOS, R. & GÓMEZ-BERBÍS, J. M. ٢٠٠٩. SPETA: Social pervasive e-Tourism advisor. *Telematics and informatics*, ٢٦, ٣١٥-٣٠٦
٢٣. GARCIA, I., SEBASTIA, L. & ONAINDIA, E. ٢٠١١. On the design of individual and group recommender systems for tourism. *Expert systems with applications*, ٣٨, ٧٦٨٣-٧٦٩٢
٢٤. GIBSON, H. & YIANNAKIS, A. ٢٠٠٢. Tourist roles: Needs and the lifecourse. *Annals of tourism research*, ٢٩, ٣٨٣-٣٥٨
٢٥. GOSAIN, A. & SARDANA, S. Handling class imbalance problem using oversampling techniques: A review. Advances in Computing, Communications and Informatics (ICACCI), ٢٠١٧ International Conference on, ٢٠١٧. IEEE, ٨٥-٧٩
٢٦. GRETZEL, U., MITSCHKE, N., HWANG, Y.-H. & FESENMAIER, D. R. ٢٠٠٤. Tell me who you are and I will tell you where to go: Use of travel personalities in destination recommendation systems. *Information Technology & Tourism*, ٧, ١٢-٣
٢٧. GRÜN, C., NEIDHARDT, J. & WERTHNER, H. ٢٠١٧. Ontology-based matchmaking to provide personalized recommendations for tourists. *Information and Communication Technologies in Tourism ٢٠١٧*. Springer.
٢٨. HAN, J., PEI, J. & KAMBER, M. ٢٠١١. *Data mining: concepts and techniques*, Elsevier.
٢٩. HUANG, Y. & BIAN, L. ٢٠٠٩. A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet. *Expert Systems with Applications*, ٣٦, ٩٤٣-٩٣٣



٣٠. KIM, K.-J. & AHN, H. ٢٠٠٨. A recommender system using GA K-means clustering in an online shopping market. *Expert systems with applications*, ٣٤, .١٢٠٩-١٢٠٠
٣١. LIANG, Y., LI, Q. & QIAN, T. Finding relevant papers based on citation relations. *International Conference on Web-Age Information Management*, ٢٠١١. Springer, -٤٠٢, ٤١٤
٣٢. LIMA, A. C. E. & DE CASTRO, L. N. ٢٠١٤. A multi-label, semi-supervised classification approach applied to personality prediction in social media. *Neural Networks*, ٥٨, .١٢٠-١٢٢
٣٣. LIU, S. M. & CHEN, J.-H. ٢٠١٥. A multi-label classification based approach for sentiment classification. *Expert Systems with Applications*, ٤٢, .١٠٩٣-١٠٨٣
٣٤. LU, J., WU, D., MAO, M., WANG, W. & ZHANG, G. ٢٠١٥ Recommender system application developments: a survey. *Decision Support Systems*, ٧٤, .٣٢-١٢
٣٥. LUCAS, J. P., LUZ, N., MORENO, M. N., ANACLETO, R., FIGUEIREDO, A. A. & MARTINS, C. ٢٠١٣. A hybrid recommendation approach for a tourism system. *Expert Systems with Applications*, ٤٠, .٣٥٥٠-٣٥٣٢
٣٦. MARTINEZ, L., RODRIGUEZ, R. M. & ESPINILLA, M. Reja: a georeferenced hybrid recommender system for restaurants. *Proceedings of the ٢٠٠٩ IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology- Volume ٠٣*, ٢٠٠٩. IEEE Computer Society, .١٩٠-١٨٧
٣٧. MORENO, A., VALLS, A., ISERN, D., MARIN, L. & BORRÀS, J. ٢٠١٣. Sigtur/e-destination: ontology-based personalized recommendation of tourism and leisure activities. *Engineering Applications of Artificial Intelligence*, ٢٦, .٦٥١-٦٣٣
٣٨. NEIDHARDT, J., SEYFANG, L., SCHUSTER, R. & WERTHNER, H. ٢٠١٥. A picture-based approach to recommender systems. *Information Technology & Tourism*, ١٥, .٦٩-٤٩
٣٩. PASHTAN, A., BLATTNER, R., ANDI, A. H. & SCHEUERMANN, P. ٢٠٠٣. CATIS: A context-aware tourist information system.
٤٠. PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R. & DUBOURG, V. ٢٠١١. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, ١٢, .٢٨٣٠-٢٨٢٥
٤١. REN, Y. ٢٠١٣. Data preprocessing for data mining.
٤٢. RICCI, F. ٢٠١١. LiorRokach, and BrachaShapira." Introduction to recommender systems handbook. springer US.
٤٣. RICCI, F., ROKACH, L. & SHAPIRA, B. ٢٠١١. Introduction to recommender systems handbook. *Recommender systems handbook*. Springer.
٤٤. RICCI, F., ROKACH, L. & SHAPIRA, B. ٢٠١٥. Recommender systems: introduction and challenges. *Recommender systems handbook*. Springer.
٤٥. RIVOLLI, A., PARKER, L. C. & DE CARVALHO, A. C. Food truck recommendation using multi-label classification. *Portuguese Conference on Artificial Intelligence*, ٢٠١٧. Springer, .٥٩٦-٥٨٥
٤٦. RUOTSALO, T., HAAV, K., STOYANOV, A., ROCHE, S., FANI, E., DELIAI, R., MÄKELÄ, E., KAUPPINEN, T. & HYVÖNEN, E. ٢٠١٣. SMARTMUSEUM: A mobile recommender system for the Web of Data. *Web semantics: Science, services and agents on the world wide web*, ٢٠, .٦٧-٥٠
٤٧. SCHAPIRE, R. ٢٠٠٨. Cos ٥١١: Theoretical machine learning. *FTP: <http://www.cs.princeton.edu/courses/archive/spr٠٨/cos٥١١/scribe notes/٠٢٠٤.pdf>*.

٤٨. SCHIAFFINO, S. & AMANDI, A. ٢٠٠٩. Building an expert travel agent as a software agent. *Expert Systems with Applications*, ٣٦, .١٢٩٩-١٢٩١
٤٩. SHARMA, S. & SHARMA, S. ١٩٩٦. Applied multivariate techniques.
٥٠. SHOVAL, N. ٢٠٠٠. Commodification and theming of the sacred: changing patterns of tourist consumption in the 'Holy Land'. *New forms of consumption: consumers, culture, and commodification*, .٢٦٣-٢٥١
٥١. SIEG, A., MOBASHER, B. & BURKE, R. D. ٢٠٠٧. Learning ontology-based user profiles: A semantic approach to personalized web search. *IEEE Intelligent Informatics Bulletin*, ٨, .١٨-٧
٥٢. SWAN, A. L., MOBASHERI, A., ALLAWAY, D., LIDDELL, S. & BACARDIT, J. ٢٠١٣. Application of machine learning to proteomics data: classification and biomarker identification in postgenomics biology. *Omics: a journal of integrative biology*, ١٧, -٥٩٥ .٦١٠
٥٣. TABACHNICK, B. G. & FIDELL, L. S. ٢٠٠٧. *Using multivariate statistics*, Allyn & Bacon/Pearson Education.
٥٤. THIENGBURANATHUM, P. ٢٠١٨. *An intelligent destination recommendation system for tourists*. Bournemouth University.
٥٥. TUNG, H.-W. & SOO, V.-W. A personalized restaurant recommender agent for mobile e-service. e-Technology, e-Commerce and e-Service, ٢٠٠٤. IEEE'٠٤. ٢٠٠٤ IEEE International Conference on, ٢٠٠٤. IEEE, .٢٦٢-٢٥٩
٥٦. VANDERPLAS, J. ٢٠١٦. *Python data science handbook: essential tools for working with data*, " O'Reilly Media, Inc."
٥٧. WEBER, I. & CASTILLO, C. The demographics of web search. Proceedings of the ٣٣rd international ACM SIGIR conference on Research and development in information retrieval, ٢٠١٠. ACM, .٥٣٠-٥٢٣
٥٨. WERTHNER, H., ALZUA-SORZABAL, A., CANTONI, L., DICKINGER, A., GRETZEL, U., JANNACH, D., NEIDHARDT, J., PRÖLL, B., RICCI, F. & SCAGLIONE, M. ٢٠١٥. Future research issues in IT and tourism. *Information Technology & Tourism*, ١٥, .١٥-١
٥٩. XUE, G.-R., LIN, C., YANG, Q., XI, W., ZENG, H.-J., YU, Y. & CHEN, Z. Scalable collaborative filtering using cluster-based smoothing. Proceedings of the ٢٨th annual international ACM SIGIR conference on Research and development in information retrieval, ٢٠٠٥. ACM, .١٢١-١١٤
٦٠. YANG, W.-S. & HWANG, S.-Y. ٢٠١٣. iTravel: A recommender system in mobile peer-to-peer environment. *Journal of Systems and Software*, ٨٦, .٢٠-١٢
٦١. YONG, A. G. & PEARCE, S. ٢٠١٣. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, ٩, .٩٤-٧٩
٦٢. ZENEBE, A. & NORCIO, A. F. ٢٠٠٩. Representation, similarity measures and aggregation methods using fuzzy sets for content-based recommender systems. *Fuzzy sets and systems*, ١٦٠, .٩٤-٧٦
٦٣. ZHENG, Y., MOBASHER, B. & BURKE, R. Context recommendation using multi-label classification. Proceedings of the ٢٠١٤ IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume ٠٢, ٢٠١٤. IEEE Computer Society, .٢٩٥-٢٨٨

٦٤. ZIEGLER, C.-N. & LAUSEN, G. Analyzing correlation between trust and user similarity in online communities. International Conference on Trust Management, ٢٠٠٤. Springer, ٢٦٥-٢٥١.
٦٥. ZINS, A. H. ٢٠٠٧. Exploring travel information search behavior beyond common frontiers. *Information Technology & Tourism*, ٩, ١٦٤-١٤٩

وبسایت‌ها

٦٦. Nooney, Kartik. “Deep Dive into Multi-Label Classification..! – Towards Data Science.” *Towards Data Science*, Towards Data Science, ٨ June ٢٠١٨, [towardsdatascience.com/journey-to-the-center-of-multi-label-classification-٣٨٤٤٠٢٢٩bfff](https://towardsdatascience.com/journey-to-the-center-of-multi-label-classification-٣٨٤٤٠٢٢٩bfff).
٦٧. Jain, Shubham. “Solving Multi-Label Classification Problems (Case Studies Included).” *Analytics Vidhya*, ٢٨ Aug. ٢٠١٧, [www.analyticsvidhya.com/blog/٢٠١٧/٠٨/introduction-to-multi-label-classification/](https://www.analyticsvidhya.com/blog/٢٠١٧/٠٨/introduction-to-multi-label-classification/).

## پیوست‌ها

## پیوست ((الف)) پرسشنامه اول

این پرسشنامه جهت سنجش علاقه مردم استان تهران به مفاهیم و فعالیت های گردشگری است که نتیجه حاصل منجر به ساخت یک الگوریتم کامپیوتری پیشنهادگر خواهد شد.

لطفا به تمام سوالات پاسخ دهید.

تصور کنید که زمان مناسب در اختیار دارید و مایل به گردش و تفریح در تهران هستید، اما قادر به بیان و مشخص کردن خواسته های خود نیستید. در زیر تعدادی فعالیت و مقصد گردشگری را مشاهده می کنید با توجه به علاقه خود به هر کدام از آنها از یک تا پنج نمره دهید. (به تمامی گزینه ها امتیاز دهید)

عدد یک: کمترین میزان علاقه عدد پنج: بیشترین میزان علاقه

○	○	○	○	○	بازدید از کاخ های تاریخی: مانند کاخ مرمر، کاخ موزه سعدآباد، کاخ نیاوران و ...
۱	۲	۳	۴	۵	
○	○	○	○	○	بازدید از بازارهای سنتی: مانند بازدید از بازار تجریش
۱	۲	۳	۴	۵	
○	○	○	○	○	بازدید از موزه های تاریخ: مانند موزه ملی، موزه هنرهای معاصر و ...
۱	۲	۳	۴	۵	
○	○	○	○	○	بازدید از بناها و عمارت ها: مانند بازدید از عمارت باغ فردوس، عمارت کلاه فرنگی و ...
۱	۲	۳	۴	۵	
○	○	○	○	○	تماشای فیلم در سینما
۱	۲	۳	۴	۵	
○	○	○	○	○	تئاتر: رفتن به تماشای تئاتر، مانند سالن تئاتر شهر
۱	۲	۳	۴	۵	
○	○	○	○	○	موزه های فرهنگی و هنری: مانند موزه فرش، موزه قرآن کریم و ..
۱	۲	۳	۴	۵	
○	○	○	○	○	بازدید از امام زاده ها: مانند امام زاده صالح، امام زاده داود و ...
۱	۲	۳	۴	۵	
○	○	○	○	○	مساجد و حسینیه های تاریخی: مانند مسجد ارگ، مسجد شاه و ...
۱	۲	۳	۴	۵	
○	○	○	○	○	رویدادهای ورزشی: مانند مسابقه فوتبال، والیبال و ..
۱	۲	۳	۴	۵	
○	○	○	○	○	ورزش های آبی: مانند شنا، قایق سواری و ...
۱	۲	۳	۴	۵	
○	○	○	○	○	ورزش های غیر آبی: مانند فوتبال، بدنسازی، والیبال و پینت بال و ...
۱	۲	۳	۴	۵	
○	○	○	○	○	نمایشگاه های هنری: مانند بازدید از یک گالری نقاشی، نگارخانه ها و ....
۱	۲	۳	۴	۵	
○	○	○	○	○	نمایشگاه های کتاب: مانند بازدید از نمایشگاه های دائمی و غیر دائمی کتاب و تعامل با افراد علاقه مند به کتاب
۱	۲	۳	۴	۵	
○	○	○	○	○	کنسرت موسیقی: رفتن به تماشای یک کنسرت موسیقی سنتی، کلاسیک، پاپ و ...
۱	۲	۳	۴	۵	
○	○	○	○	○	مراکز خرید: گشت و گذار در مراکز بزرگ خرید
۱	۲	۳	۴	۵	
○	○	○	○	○	باغ های عمومی: مانند بازدید و گشت و گذار در باغ های پرندگان، باغ نگارستان، باغ موزه قصر و ...
۱	۲	۳	۴	۵	
○	○	○	○	○	رستوران ها: صرف شام یا ناهار در یک رستوران
۱	۲	۳	۴	۵	
○	○	○	○	○	کافه: رفتن به کافه ها و گپ زدن با دوستان و آشنایان
۱	۲	۳	۴	۵	
○	○	○	○	○	باغ وحش: بازدید از باغ وحش
۱	۲	۳	۴	۵	

بوستان ها: گردش و تفریح در بوستان های مانند بوستان آب و آتش، بوستان نهج البلاغه و ...

○	○	○	○	○	روستاها: بازدید و سیاحت در روستاهای خوش آب و هوا
۱	۲	۳	۴	۵	
○	○	○	○	○	دریاچه ها: گردش و سفر در کنار دریاچه هایی مانند دریاچه چیتگر
۱	۲	۳	۴	۵	
○	○	○	○	○	آبشارها و رودها: مانند آبشار اوسند دریند، مانند رودخانه کرج
۱	۲	۳	۴	۵	
○	○	○	○	○	کوه ها: پیاده روی و کوهنوردی در کوههایی مانند درکه، توچال و ...
۱	۲	۳	۴	۵	

پایان

سپاس از همکاری و بردباری شما

## پیوست ((ب)) پرسشنامه دوم

این پرسشنامه جهت سنجش ارتباط دسته‌های گردشگری و علاقه به مفاهیم و فعالیت های گردشگری است که نتیجه حاصل منجر به ساخت یک الگوریتم کامپیوتری پیشنهادگر خواهد شد.  
لطفا به تمام سوالات پاسخ دهید.

با توجه به نزدیکی هر کدام از گروه‌های گردشگری به علائق شما، امتیاز یک تا پنج را انتخاب کنید.(به تمامی سوالات پاسخ دهید)  
عدد یک: کمترین ارتباط    عدد پنج: بیشترین ارتباط

تاریخی: شما به تاریخ، بناهای و آثار تاریخی علاقه‌مند هستید.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
۱	۲	۳	۴	۵

مفرح: شما اهل قرار گذاشتن با دوستان، تفریحات کوتاه روزانه هستید.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
۱	۲	۳	۴	۵

طبیعت گرد: شما عاشق سکوت و طبیعت هستید و با بودن در طبیعت بودن احساس آرامش می‌کنید.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
۱	۲	۳	۴	۵

ورزشی: شما بخشی از انرژی و زمان خود را برای ورزش و مسائل مربوط به ورزش می‌گذارید و به طور مرتب آن را دنبال می‌کنید.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
۱	۲	۳	۴	۵

فرهنگی و هنری: اخبار و مسائل مربوط به فرهنگ و هنر را دنبال می‌کنید و در صورت امکان اوقات خود را با این دست فعالیت‌ها می‌گذرانید.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
۱	۲	۳	۴	۵

مذهبی: شما به معنویات و تعالی روحی خود توجه دارید و سعی می‌کنید بخشی از اوقات خود را به امور مذهبی بپردازید.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
۱	۲	۳	۴	۵

طبیعت گرد شهری: شما به طبیعت علاقه دارید، اما در زمانی که امکان دوری از شهر نیست، از طبیعت‌های شهری برای اوقات فراغت خود نیز

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
۱	۲	۳	۴	۵

استفاده می‌کنید.

### دسته دوم سوالات گردشگری

تصور کنید که زمان مناسب در اختیار دارید و مایل به گردش و تفریح در تهران هستید، اما قادر به بیان و مشخص کردن خواسته های خود نیستید.  
در زیر تعدادی فعالیت و مقصد گردشگری را مشاهده می‌کنید با توجه به علاقه خود هر تعداد فعالیت را که مایل بودید انتخاب کنید.

بازدید از کاخ های تاریخی، بناها و عمارت ها: مانند کاخ مرمر، کاخ موزه سعدآباد، کاخ نیاوران، عمارت باغ فردوس، و ...    ☐ بلی    ☐ خیر

بازدید از بازارهای سنتی و گشت و گذار در مراکز خرید: مانند بازدید از بازار تجریش    ☐ بلی    ☐ خیر

بازدید از موزه های تاریخ و فرهنگ: مانند موزه ملی، موزه هنرهای معاصر و ...	<input type="radio"/> بلی	<input type="radio"/> خیر
تماشای فیلم در سینما	<input type="radio"/> بلی	<input type="radio"/> خیر
تئاتر: رفتن به تماشای تئاتر، مانند سالن تئاتر شهر	<input type="radio"/> بلی	<input type="radio"/> خیر
بازدید از امام زاده ها: مانند امام زاده صالح، امام زاده داود و ...	<input type="radio"/> بلی	<input type="radio"/> خیر
مساجد و حسینیه های تاریخی: مانند مسجد ارگ، مسجد شاه و ...	<input type="radio"/> بلی	<input type="radio"/> خیر
رویدادهای ورزشی: مانند مسابقه فوتبال، والیبال و ....	<input type="radio"/> بلی	<input type="radio"/> خیر
ورزش های آبی و غیر آبی: مانند شنا، قایق سواری، فوتبال، بدنسازی و ...	<input type="radio"/> بلی	<input type="radio"/> خیر
نمایشگاه های و رویدادهای هنری و فرهنگی: مانند بازدید از یک گالری نقاشی، نگارخانه ها، بازدید از نمایشگاه های دائمی و غیر دائمی کتاب....	<input type="radio"/> بلی	<input type="radio"/> خیر
باغ های و بوستان های عمومی: مانند بازدید و گشت و گذار در باغ های پرندگان، باغ نگارستان، باغ موزه قصر، بوستان آب و آتش و ...	<input type="radio"/> بلی	<input type="radio"/> خیر
رستوران ها: صرف شام یا ناهار در یک رستوران	<input type="radio"/> بلی	<input type="radio"/> خیر
کافه: رفتن به کافه ها و گپ زدن با دوستان و آشنایان	<input type="radio"/> بلی	<input type="radio"/> خیر
باغ وحش: مانند باغ وحش ارم	<input type="radio"/> بلی	<input type="radio"/> خیر
روستاها: بازدید و سیاحت در روستاهای خوش آب و هوا	<input type="radio"/> بلی	<input type="radio"/> خیر
آبشارها و رودها و دریاچه ها: مانند آبشار اوسند دربند، مانند رودخانه کرج، دریاچه چیتگر و ...	<input type="radio"/> بلی	<input type="radio"/> خیر
کوه ها: پیاده روی و کوهنوردی در کوه هایی مانند درکه، توچال و ...	<input type="radio"/> بلی	<input type="radio"/> خیر

پایان

سپاس از همکاری و بردباری شما





## **Abstract**

Tourism and leisure play a very important role in human life, and in the meantime, information technology in combination with the tourism industry has been a twofold advantage for users and providers of leisure and tourism services. On the other hand, technology has increased the excessive amount of information on the web, which leads to redundancy of information and thus the user is in trouble to get the right information for decision-making, and the lack of domain knowledge makes the user unable to express his wishes clearly to search engines. One solution to overcome this problem is using recommender systems that use different techniques and approaches. Ontology-based recommender systems require experts to judge the relationships of elements in ontology levels. On the other hand, common approaches such as clustering may not be suitable for any given data and do not provide appropriate clusters, as well as the proposed recommendations are not appropriately diversified; on the other hand, to resolve the problem of cold start, the reception of explicit demographic information from the user can be unpleasant for users. In this research, Tehran province has been selected as a case study, second-level groups of ontology have been obtained through factorial analysis and a method has been developed which, using a multilabel classification approach, establishes the relationship between the level of groups and the level of ontology concept without human judgment. In the method presented in this study, information received from the user does not include demographic data, and only the user's interest in tourism groups is measured. The results evaluated by an offline evaluation, scenario-based and user study methods, and in all cases responded to the research objectives.

**Keys:** recommender systems, tourism, multilabel classification, ontology, factorial analysis, stereotypes



**Tarbiat Modares University**  
**Faculty of Engineering**  
**Information Technology Engineering Department**

**Thesis report**

**Designing a recommender system for  
tourist sites  
(Case study: Tehran province)**

**Student**  
**Rasoul norouzi**  
**Supervisor**  
**Dr. Amir Albadvi**  
**Advisor**  
**Dr. Elham Akhondzadeh**

**September 2018**