بسمه تعالی

فرم معرفی پروژه کارشناسی

پردیس دانشکدگان فنی- دانشکده علوم مهندسی

1- عنوان پروژه کارشناسی: Frequent Pattern Mining

2- مشخصات دانشجو

نام و نام خانوادگی: مینو احمدی　　شماره دانشجویی: 810897032　　خوشه: بهینه سازی

3- استاد راهنما : دکتر علی فهیم

4- نیمسال اخذ واحد پروژه: ترم نیم سال دوم سال ۱۴۰۱-۱۴۰۰

5- اطلاعات مربوط به پروژه

الف – تعریف مسئله:

در شرکت‌های صنعتی، یکی از راه های مدرن بازاریابی و افزایش فروش محصولات، استفاده از الگوریتم های داده کاوی برای کشف الگو های خرید و فروش های انجام شده و رابطه ی بین کالا ها است. با کشف این الگو ها و روابط میتوانیم به خریداران، کالاهایی که به احتمال زیاد نیاز دارند ولی نخریده اند را معرفی کرده و کالا هایی که قبلا تهیه کرده اند ولی مدتی است که خرید مجدد نکرده اند را با آفرهای هوشمندانه پیشنهاد کنیم. از این طریق با جلب نظر مشتریان به کمک متد های بازار یابی (تخفیف و پیشنهاد های ویژه) فروش شرکت بالا رفته و هزینه های بازاریابی شرکت کاهش میابد. هدف پروژه نوشتن یک اپلیکیشن تحت وب است که از طریق آن شرکت‌ها بتوانند از این امکانات در فروش خود استفاده کنند.

اغلب الگوریتم های یادگیری ماشین در داده کاوی با داده های عددی کار میکنند و در پیاده سازی و نحوه کار آنها گرایش به ریاضیات محض وجود دارد. اما، کاوش قواعد وابستگی (association rule mining)که از آن بi عنوان کاوش قواعد انجمنی نیز یاد میشود، برای داده های دسته ای مناسب و محاسبات آن نسبت به بسیاری از دیگر الگوریتم ها ساده تر است. این روش، یکی از راهکارهای مبتنی بر قواعد (rules)، برای کشف روابط جالب بین متغیرها در پایگاه داده های بزرگ محسوب میشود. در کاوش قواعد وابستگی، قواعد قوی با استفاده از سنجه جذابیت (interestingness)شناسایی میشوند .

مساله کاوش قواعد وابستگی را میتوان به صورت ریاضی و چنانچه در ادامه می آید دید .

- I = {i1, i2, ...} مجموعه ای از ویژگی های دودویی است که به آنها اقلام گفته میشود.
- D = {t1, t2, ...} مجموعهای از تراکنشها است که پایگاه داده نامیده میشود .
- هر تراکنش در D شامل زیرمجموعهای از اقلام موجود در I است .

قواعد ساده وابستگی / انجمنی به صورت زیر هستند. لازم به ذکر است که در قاعده زیر، t1مقدم و t2نتیجه (موخر) محسوب میشود .

- t1 ⇒ t2 (در اینجا، ti به طور کلی یک مورد مجزا یا مجموعه ای از اقلام است) .

در داده کاوی و به ویژه کاوش قواعد وابستگی، به منظور انتخاب قواعد جالب از میان مجموعه ای از قواعد ممکن، محدودیت های گوناگونی (به عنوان آستانه) بر سنجه های مختلف مرتبط با اهمیت و جالبی (interestingness)، اعمال میشود . شناخته شده ترین محدودیت ها در کاوش قواعد وابستگی، آستانه کمینه برای پشتیبان (support)و اطمینان (confidence) هستند. همچنین، معیارهای دیگری از جمله بالابری (Lift)و عقیده (Conviction)نیز در همین راستا مورد استفاده قرار میگیرند .

پشتیبان (Support)

پشتیبان شاخصی است از اینکه یک مجموعه اقلام (itemset)چند بار در یک مجموعه داده (data set)ظاهر میشود .

پشتیبان X، با توجه به مجموعه تراکنش T، به صورت کسر تراکنش های tدر مجموعه دادهای که شامل مجموعه اقلام X است تعریف میشود .درواقع پشتیبان Xمعادل است با کسر تعداد دفعاتی که قلم جنس Xدر تراکنشها (سفارشات) حاضر شده است .

$$\mathrm{supp}(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|}$$

اطمینان (Confidence)

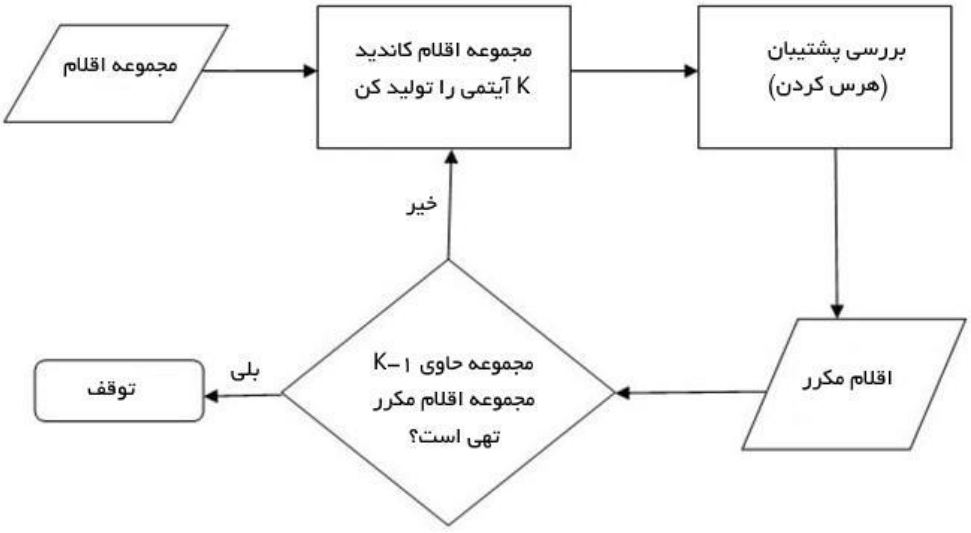
اطمینان شاخصی است از اینکه یک قاعده چند بار درست (True)بوده. مقدار اطمینان یک قاعده (X→Y)، با توجه به مجموعه تراکنش T، عبارت است از کسری از تراکنشهای شامل X که شامل Y نیز هستند. اطمینان به صورت زیر تعریف میشود .

$$\mathrm{conf}(X \Rightarrow Y) = \mathrm{supp}(X \cup Y)/\mathrm{supp}(X)$$

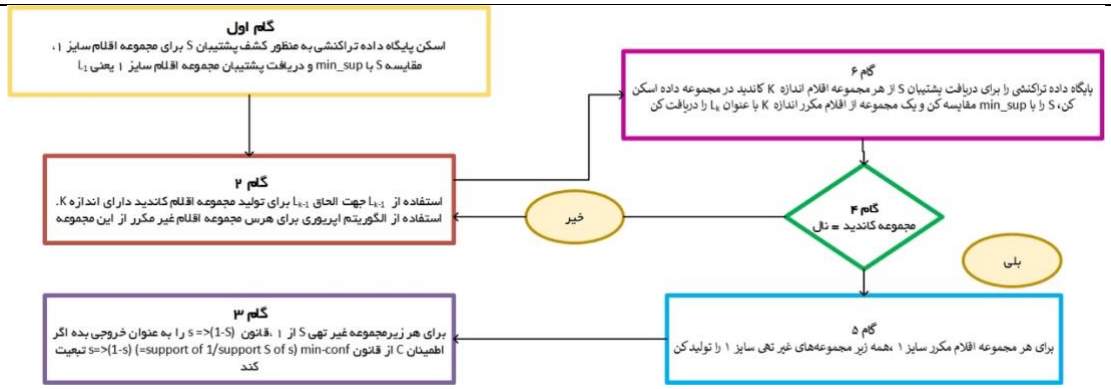
مفهوم مطرح شده کمی گیج کننده است زیرا معمولا به طور طبیعی به احتمال رویدادها و نه مجموعه اقلام فکر میشود. میتوان sup(XUY) را به صورت احتمال P(EX∩EY)بازنویسی کرد که در آن Exو Eyرویدادهایی هستند که تراکنش در آنها به ترتیب شامل مجموعه اقلام Xو Yاست. اطمینان را میتوان تخمین احتمال شرطی P(EX|EY)تفسیر کرد که احتمال یافتن RHS قواعد در تراکنش های تحت این شرط وجود دارد، که آنها نیز شامل LHS باشند .

الگوریتم Apriori

الگوریتم اپریوری (Apriori)بر این اصل بنا شده که اگر یک مجموعه اقلام (itemset) مکرر است، پس همه زیرمجموعه های آن نیز مکرر هستند. این بدین معنا است که اگر {۰،۱} مکرر باشد، پس {۰} و {۱} نیز مکرر هستند. بالعکس این قاعده نیز صادق است، یعنی اگر یک مجموعه اقلام مکرر نباشد، زیرمجموعه های آن نیز مکرر نیستند. با توجه به توضیحات بالا، برای یافتن یک مجموعه قواعد وابستگی، ابتدا باید مجموعه اقلام مکرر را پیدا کرد. برای حل این مساله، نیاز به کار با نوع داده های عددی و اسمی (دستهای) است .



برای کسب درک بهتر از الگوریتم میتوان برخی کاربردهای آن مانند «تحلیل سبد خرید» را مورد بررسی قرار داد. در این کاربرد، داده کاو به دنبال کشف آن است که کدام اقلام با یکدیگر (در یک سبد خرید) خریداری شده اند. در دیگر مثالی که میتوان پیرامون الگوهای مکرر زد، ابزارهای تحلیل مالی هستند که با بهرهگیری از الگوریتم اپریوری چگونگی داغ شدن سهامهای گوناگون با یکدیگر را نمایش میدهند. فلوچارت الگوریتم اپریوری (Apriori)در ادامه آورده شده است .



این روش، ممکن است به طور پیوسته با دیگر الگوریتم ها استفاده شود تا به طور موثری داده ها را مرتب سازی و با یکدیگر مقایسه کند .اصل اپریوری میتواند تعداد اقلامی که نیاز به بررسی آنها است را کاهش دهد. این روش بیان میکند که اگر یک مجموعه اقلام فاقد تکرار است، پس همه زیرمجموعه های آن نیز نادر هستند. این امر بدین معناست که اگر {آبجو} فاقد تکرار بود، میتوان انتظار داشت که {آبجو، پیتزا} هم به همان میزان و یا حتی بیشتر، نادر باشند. بنابراین، برای یکی کردن لیست مجموعه اقلام محبوب، نیازی به در نظر گرفتن {آبجو، پیتزا} و یا هیچ یک از دیگر مجموعه اقلام حاوی آبجو، نخواهد بود .

محدودیت ها

هزینه محاسباتی بالا :روش اپریوری به لحاظ محاسباتی بسیار پر هزینه است. حتی اگر الگوریتم اپریوری تعداد اقلام کاندید برای بررسی را کاهش دهد، در صورتی که موجودی فروشگاه زیاد یا آستانه پشتیبان کم باشد میزان باقیمانده همچنان عدد بزرگی خواهد بود. یک راهکار جایگزین، کاهش تعداد مقایسهها با استفاده از ساختارهای پیشرفته داده مانند جدولهای هش برای مرتبسازی اقلام کاندید به شیوه موثرتر است .

انجمنهای جعلی :تحلیل صورت کالاهای بزرگ مجموعه اقلام بیشتری را در برمیگیرد و آستانه پشتیبان ممکن است برای شناسایی انجمنهای مشخصی کاهش پیدا کند. اگرچه، کاهش آستانه پشتیبان ممکن است تعداد انجمنهای جعلی کشف شده را

افزایش دهد. برای حصول اطمینان از اینکه انجمنهای شناسایی شده قابل تعمیم هستند، میتوان آنها را از مجموعه داده آموزش به دست آورد، پیش از آنکه پشتیبان و اطمینان ارزیابی شده برای آنها در یک مجموعه داده جدا قرا گیرد .

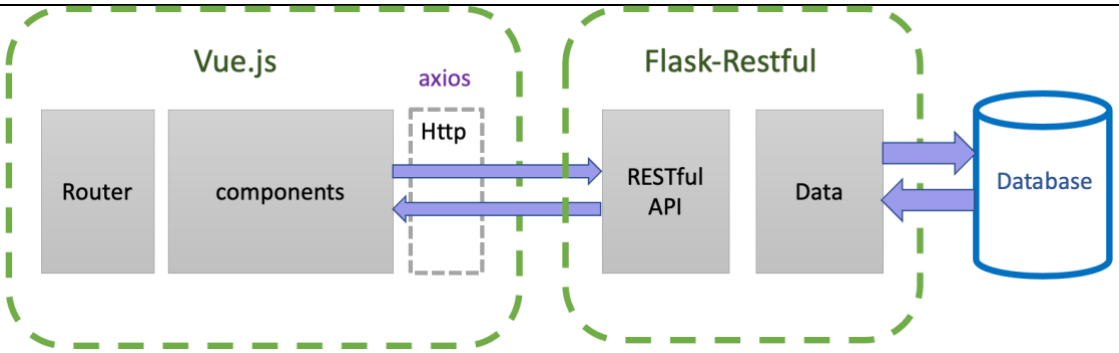
ب – هدف از طرح مورد نظر و ضرورت انجام آن:

شرکت با داشتن این قابلیت‌ها برای درآمد یکسان، برای بازاریابی مشتریان جدید هزینه کمتری می‌نماید زیرا هزینه نگه داشتن مشتری قدیمی کمتر از هزینه جذب مشتری جدید است. این قابلیت باعث شناسایی هوشمند و دقیق کالاهای مورد نیاز مشتریان شده و باعث بالا رفتن وفاداری مشتریان خواهد شد.

ج- روش‌های اجرایی انجام پروژه:

در حالت کلی ابتدا به دنبال داده های مناسب میگردم که بتوان با آن نزدیکترین شبیه سازی را با داده های واقعی کرد. سپس به پیاده سازی الگوریتم آن با استفاده از الگوریتم های داده کاوی می پردازیم. بعد از آماده شدن الگوریتم به درست کردن بخش سرور اپلیکیشن می پردازیم و در ادامه آن به درست کردن یک صفحه که بتوان این کارها را در آن نشان داد . برای داده ها از داده های دانشگاه UCIاستفاده میکنم. این داده ها شامل تراکنش های سفارش های یک مغازه است. سپس برای شروع از الگوریتم اپریوری که در بالا توضیح داده شد استفاده می شود که که بتوانم قواعد انجمنی را استخراج کنم و همچنین در آخر از بقیه الگوریتم ها نیز استفاده میکنم که بتوانم بهترین پاسخ را انتخاب کنم .

در ادامه به سازگار کردن این کد ها با بخش سرور نرم افزار یا همان بخش back-end با استفاده از Flask Restful که توسط اقای افضلی انجام شد .در آخراین پروژه قسمت فرانت اند را به کمکِ فریم ورک جاوا اسکریپت به نام Vue.js انجام می دهیم ، چرا که کار با این فریم ورک بسیار راحت هست و امکانات زیادی رو در اختیار ما قرار میدهد. در شکل زیر نحوه ی پیاده سازی در فرانت اند نمایش داده شده است ، در واقع این قسمت که فرانت اند هست و با ویو ایجاد میشود. کامپوننت ها (بخش های مختلف سایت مثل فوتر هدر ..) را بعد از پیاده سازی و طراحی با به کار گیری کتابخانه روتر ان ها را به هم لینک کرده و با استفاده از کتابخانه اکسیوس با بک اند ارتباط برقرار کرده و لینک میشویم.



د- برنامه زمانی:

برنامه زمانی بر اساس Gantt Chart زیر پیش خواهد رفت :

GANTT CHART TIME SCHEDULE

PROJECT	START DATE
FREQUENT PATTERN MINING	Thursday, September 23, 2021

User to complete non-shaded fields only.

TASKS	TASK OWNER	STATUS
Phase 1: Algorithms		
Reading about algorithms	Aref	Complete
Generating data	Aref	Canceled
Finding data	Aref	Complete
Finding rules	Aref	Complete
Suggest items based on basket	Aref	Complete
Suggest customer's ID based on consequences of rules	Aref	Complete
Creating Graph based on rules	Aref	Needs Update
Optimizing algorithm	Aref & Minoo	In Progress
Test another algorithms	Aref & Minoo	Not Started
Phase 2: Back-End		
Learning Flask	Aref	Complete
Design database	Aref	Complete
Creating database	Aref	Not Started
Object oriented design	Aref	Needs Review
Adjusting algorithms' code base with back-end structure	Aref	Complete
Error Handling	Aref	Not Started
Creating end-points	Aref	Needs Update
Refactoring	Aref	Not Started
Phase 3: Front-End		
Learning Vue-js	Aref	Complete
searching about showing graph page	Minoo & Aref	Canceled
packaging(npm)	Minoo	In Progress
Rules page design	Minoo	In Progress
Rules page developement	Minoo	In Progress
market page design	Minoo	In Progress
market page developement	Minoo	In Progress
contact page design	Minoo	In Progress
contact page development	MInoo	Not Started
about page design	MInoo	In Progress
about page development	MInoo	Not Started
Final Phase: Report		
Desiging a poster	Aref & Minoo	Not Started
Writing report	Aref & Minoo	Not Started

STATUS
Not Started
In Progress
Complete
On Hold
Overdue
Needs Review
Needs Update
Canceled

ه- پروژه در ارتباط با کدام سازمان، واحد صنعتی، پروژه کارشناسی یا آزمایشگاه است:

واحد ها و شرکت های پخش

و- مراجع اصلی:

[1] Zhan, Foxiao & Zhu, Xiaolan & Zhang, Lei & Wang, Xuexi & Wang, Lu & Liu, Chaoyi. (2019). Summary of Association Rules. IOP Conference Series: Earth and Environmental Science. 252. .10.1088/1755-1315/252/3/032219 .032219

[2] Mohammed J. Zaki, Wagner Meira, Jr., Data Mining and Machine Learning: Fundamental Concepts and Algorithms, 2nd Edition, Cambridge University Press, March 2020. ISBN: 978- .1108473989

[3] Grinberg M. Flask web development: developing web applications with python. " O'Reilly Media, Inc." 2018.

[4] Pšenák, Peter & Tibensky, Matus. (2020). The usage of Vue JS framework for web application creation. Mesterséges intelligencia. 2. 61-72. 10.35406/MI.2020.2.61.

[5] Hwang, Y. H. (2019). Hands-on data science for marketing: Improve your marketing strategies with machine learning using Python and R.

6- <u>تاریخ و امضاء دانشجو و استاد راهنما</u>

دانشجو: تاریخ :1400/11/17

استاد راهنما: تاریخ :

7- پروژه کارشناسی آقای/خانم ..... با شماره دانشجویی ..... در تاریخ ..... داوری و با نمرات زیر مورد تصویب قرار گرفت.

مسئولیت	نمره	امضا
استاد راهنما (نمره از 15)		
نمره داوران (نمره از 5)		