|  |  |  |
| --- | --- | --- |
| دانشگاه تهران - ویکی‌پدیا، دانشنامهٔ آزاد | دانشگاه تهران  دانشگدگان فنی  دانشکده‌ی علوم مهندسی | Description: Related image |
|  | | |
|  | | |
| Frequent Pattern Mining  پروژه براي دريافت درجه کارشناسی  در رشته‌ی علوم مهندسی | | |
| **مینو احمدی** | | |
| استاد راهنما:  دکتر علی فهیم | | |
| **بهمن ماه 1400** | | |

**تقديم به تمام کسانی که در راه علم اموزی در طول زندگی مرا یاری نموده اند.**

**-با تشکر و سپاس از استاد دانشمند و پر مایه ام جناب آقای دکتر علی فهیم که از محضر پر فیض تدریسشان، بهره ‌ها برده‌ام.**

**-با امتنان بیکران از مساعدت های بی شائبه ی جناب آقای دکتر معینی ریاست محترم دانشکده علوم مهندسی دانشگاه تهران**

**-با سپاس بی‌دریغ خدمت دوستان گران‌ مایه ‌ام اقای عارف افضلی که مرا صمیمانه و مشفقانه یاری داده‌اند.**

**چکيده**

در این پروژه هدف راه‌اندازی نرم‌افزاری برای بهینه‌سازی و بهبود فرآیند فروش شرکت‌ها است. ما در این سیستم با استفاده از الگوریتم‌های داده‌کاوی و ایجاد قوانینی مبنی بر فاکتورهای شرکت، تلاش کردیم پیشنهادهایی کاربردی و دقیق به صاحبان کسب‌وکار بدهیم تا آن‌ها بتوانند با استفاده از این پیشنهادها بازاریابی خود را دقیق‌تر و کارآمدتر کنند. این پیشنهادها شامل دو بخش است. اولین نوع پیشنهاد بر پایه سبد خرید فعلی مشتری صورت می‌گیرد که کالاهای مرتبط با سبد او را به کارفرما می‌دهد تا کارفرما بتواند با پیشنهاد آن اجناس، احتمال فروششان را بالا ببرد. نوع دوم پیشنهاد در سیستم بر پایه خریدهای قبلی مشتریان است که به کارفرما بر اساس کالاهای انتخابی‌اش، مشتری‌های احتمالی‌ای که خواهان آن اجناس هستند را پیشنهاد می‌کند. لازم به ذکر است که با توجه به سخت بودن استفاده از الگوریتم‌ها برای شرکت‌ها ما تلاش کردیم تا برای اولین بار در ایران بستری را فراهم سازیم تا شرکت‌ها با کمک یک نرم‌افزار به راحتی بتوانند با چند کلیک ساده نتایج را استخراج کنند، در واقع با پیاده‌سازی الگوریتم‌های پیچیده در بستری که استفاده آن برای کاربران بسیار آسان است، دسترسی شرکت‌ها را به این نتایج و پیشنهادات ممکن ساختیم.

در ساخت این نرم‌افزار از ابزار Flask restful برای توسعه‌ی back-end استفاده شده است. همچنین برای توسعه front-end از ابزار Vue-JS استفاده شده است. داده‌های کل سیستم در پایگاه‌داده‌ای با سیستم مدیریتی postgresql ذخیره می‌شود. در هسته این سیستم نیز از الگوریتم‌های Apriori و FP-Growth استفاده شده است که به نقاط مثبت و منفی آن‌ها در گزارش اشاره شده است.

**واژه‌هاي كليدي**: داده‌کاوی – قوائد انجمنی – Frequent Pattern Mining – آنالیز سبد خرید

فهرست مطالب

[فصل 1: مقدمه 1](#_Toc95090693)

[1-1- مقدمه 1](#_Toc95090694)

[1-1-1- عنوان تحقیق 2](#_Toc95090695)

[1-1-2- مسئله تحقیق 3](#_Toc95090696)

[1-1-3- تاریخچه ای از موضوع تحقیق 3](#_Toc95090697)

[1-1-4- تعریف موضوع تحقیق 5](#_Toc95090698)

[1-1-5- هدف یا هدف‌های کلی و آرمانی تحقیق 5](#_Toc95090699)

[1-1-6- روش انجام پروژه 6](#_Toc95090700)

[1-1-7- نوآوری، اهمیت و ارزش تحقیق 6](#_Toc95090701)

[1-1-8- خلاصه فصل‌ها 7](#_Toc95090702)

[فصل 2: مروري بر مطالعات انجام شده 8](#_Toc95090703)

[2-1- مقدمه 8](#_Toc95090704)

[2-2- تعاريف، اصول و مباني نظري 9](#_Toc95090705)

[پشتیبان (Support) 9](#_Toc95090706)

[اطمینان (Confidence) 9](#_Toc95090707)

[2-3- مروري بر ادبيات موضوع 10](#_Toc95090708)

[الگوریتم Apriori 10](#_Toc95090709)

[الگوریتم FP-Growth 12](#_Toc95090710)

[2-4- نتيجه‌گيري 13](#_Toc95090711)

[محدودیت‌های الگوریتم Apriori 13](#_Toc95090712)

[Apriori vs FP-Growth 13](#_Toc95090713)

[فصل 3: روش تحقيق 15](#_Toc95090714)

[3-1- مقدمه 15](#_Toc95090715)

[3-2- الگوکاوی فراوانی 15](#_Toc95090716)

[3-2-1- علت انتخاب روش 15](#_Toc95090717)

[3-2-2- تشريح كامل روش تحقيق 16](#_Toc95090718)

[فصل 4: نتايج 20](#_Toc95090719)

[4-1- مقدمه 20](#_Toc95090720)

[4-2- محتوا 20](#_Toc95090721)

[4-3- اعتبار سنجی 27](#_Toc95090722)

[فصل 5: بحث و نتیجه گیری 28](#_Toc95090723)

[5-1- مقدمه 28](#_Toc95090724)

[5-2- محتوا 29](#_Toc95090725)

[5-2-1- جمع‌بندي 29](#_Toc95090726)

[5-2-2- نوآوري 30](#_Toc95090727)

[5-2-3- پيشنهادها 30](#_Toc95090728)

[5-2-4- محدودیتها 30](#_Toc95090729)

[فصل 6: مراجع 32](#_Toc95090730)

**فهرست اشکال**

[شکل 1- الگوریتم اپریوری۱ 9](#_Toc108184787)

[شکل 2-الگوریتم اپریوری ۲ 9](#_Toc108184788)

[شکل 3-الگوریتم FP-Growth 11](#_Toc108184789)

[شکل 4- تفاوت الگوریتم اپریوری و FP-Growth 12](#_Toc108184790)

[شکل 5– نمودار پکیج‌ها 15](#_Toc108184791)

[شکل 6-نمودار طراحی شی‌گرا 16](#_Toc108184792)

[شکل 7- بخش های مختلف ویو 17](#_Toc108184793)

[شکل 8-نحوه ارتباط front-end و back-end 17](#_Toc108184794)

[شکل 9 - لیست کشورها 18](#_Toc108184795)

[شکل 10- انتخاب کشور مورد نظر 19](#_Toc108184796)

[شکل 11– لیست محصولات 19](#_Toc108184797)

[شکل 12- پیشنهاد محصول 20](#_Toc108184798)

[شکل 13– پیشنهاد مشتری 21](#_Toc108184799)

[شکل 14– نمایش خطا 21](#_Toc108184800)

[شکل 15– صفحه گرفتن قوائد 22](#_Toc108184801)

[شکل 16- صفحه پیشنهادات 23](#_Toc108184802)

[شکل 17– صفحه درباره ما 23](#_Toc108184803)

[شکل 18- صفحه تماس با ما 24](#_Toc108184804)

[شکل 19- مقایسه الگوریتم‌های پروژه 26](#_Toc108184805)

1. مقدمه
   1. مقدمه

امروزه تاثیر بازاریابی در میزان افزایش فروش و سود شرکت‌ها بر همگان واضح و روشن است. ما در این تحقیق تلاش کرده‌ایم تا با استفاده از علم داده‌کاوی و با بهره‌گیری از قواعد انجمنی سبد خرید مشتریان را تحلیل کنیم. تحلیل سبد بازار، یک تکنیک مدل سازی است که بر مبنای این تئوری پایه ریزی شده است که اگر شما یک گروه ویژه و خاصی از اقلام را خریداری کنید، می‌توان گفت احتمال بیشتری وجود دارد که گروه دیگری از اقلام را نیز خریداری کنید. این تکنیک به شرکت‌ها اجازه می‌دهد تا رفتار خرید یک خریدار را درک کند و به آن‌ها کمک کند تا نیازهای خریدار را بشناسد و پیشنهادهای جذابی را برای خرید آن محصولات به مشتریان بدهد و یا محصولاتی را به فروشگاه اضافه کرده و یا چیدمان فروشگاه را مطابق با آن اطلاعات تغییر دهد. ما در این تحقیق برای تحلیل درست و دقیق سبد بازار الگوریتم‌های مختلفی را به کار برده و به مقایسه معایب و مزایای آن‌ها پرداخته‌ایم تا در نهایت بهترین الگوریتم را به کار بگیریم. لازم به ذکر است که نتایج این تحقیق تماما به صورت عملی به صورت یک وب اپلیکشین پیاده سازی شده است و بر روی داده‌های واقعی شرکت‌های بزرگ و معروف پیاده سازی شده و نتایج آن بدست امده است.

این تحقیق منجر به طراحی و ساخت وب اپلیکیشن Basket analysis شد که تمام شرکت‌ها با اتصال پایگاه داده‌ی خود به اپلیکشین می‌توانند پیشنهادات موثر را کشف کنند و از نتایج حاصل برای بازاریابی هرچه بهتر و با هزینه کمتر استفاده کنند. پیاده سازی تمام قسمت‌های این وب اپلیکشین اعم از: ساخت و طراحی فرانت-اند، ساخت و طراحی بک-اند، پیاده‌سازی تمام الگوریتم‌ها و ... اپلیکشن تماما به دست محققان این تحقیق صورت گرفته است.

در نتیجه شرکت با استفاده از این وب اپلیکیشن 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 بالا رفتن وفاداری مشتری می‌شود. با معرفی مناسب محصولات، فضای رشد محصولات جدید پر رونق‌تر می‌شود و این مساله به بازاریابی تولیدی‌های نوپا یا محصولات کمتر شناخته شده در کسری از جامعه کمک می‌کند.

* + 1. عنوان تحقیق

تحلیل سبد خرید یکی از مهم­ترین کاربردهای داده­کاوی است که بر کشف الگوهای خرید به وسیله داده­های معاملات انجام شده، تمرکز می­کند. در بسیاری از فروشگاه­ها تبادلات، تنها منبع اطلاعات فروش هستند که داده­کاوی می­تواند از آن استفاده کند. داده­کاوی به علم استخراج دانش از داده­های خام گفته می­شود و به وسیله کاوش داده­های موجود در بانک‌های اطلاعاتی، اطلاعات و دانش مفید از داده­ها را استخراج می­کند. داده­کاوی در کاربردهای متنوعی مورد استفاده قرار­می­گیرد و یکی از ابزارهای مدیریت ارتباط با مشتری می‌باشد. با توجه به اهمیت به کارگیری تکنیک­های داده­کاوی در استخراج دانش از حجم عظیم داده­های فروش و تأثیر استفاده از این اطلاعات در سودآوری و بهبود سیستم مدیریت ارتباط با مشتری در این مقاله سعی بر آن است با استفاده از تکنیک­های داده­کاوی به تحلیل سبد خرید مشتریان یکی از شعب فروشگاه زنجیره­ای در شهر اراک بپردازیم. نتایج این مقاله نشان می­دهد به کارگیری تکنیک­های داده­کاوی در کشف اقلامی که با هم خریداری می­شوند باعث افزایش رضایت مشتریان و به دنبال آن سود فروشگاه می­شود. در این راستا از الگوریتم اجتماع جهت کشف الگوی خرید مشتریان استفاده­ شده­است. در نهایت راهکارهایی جهت بهبود عملکرد فروشگاه ارائه خواهد شد.

* + 1. مسئله تحقیق

در شرکت‌های صنعتی، یکی از راه‌های بهتر کردن فروش محصولات این است که رابطه بین کالاها را بدست بیاوریم و از طریق آن رابطه‌ها می‌توانیم به خریدارها کالاهایی که در نظر نگرفته‌اند ولی به احتمال زیادی به آن ها نیاز دارند را معرفی کنیم. از این طریق هم فروش شرکت بالا می‌رود و هم با دادن تخفیف یا پیشنهادهای ویژه به خریدار، نظر وی را جلب می‌کنیم. از طرفی از طریق این روابط می‌توانیم بر اساس خریدهای قبلی افراد، افرادی را برای خریداری کالاهایی که مثلا در انبار مانده‌اند یا کارفرما می خواهد زودتر بفروش بروند، پیشنهاد کنیم. هدف پروژه نوشتن یک اپلیکیشن تحت وب است که از طریق آن شرکت‌ها بتوانند از این امکانات در فروش خود استفاده کنند.

* + 1. تاریخچه ای از موضوع تحقیق

راماجري و دسي (2013)، در تحقيقي تحت عنوان "نقش داده‌كاوي در بخش خرده‌فروشي"، ضمن بيان مفهوم داده‌كاوي به عنوان يك فن، به كشف الگوها براي بهبود رقابت در بازار به ويژه در صنعت خرده فروشي تأكيد مي‌كنند. A.A. و همكاران (۲۰۱۲) ، در پژوهشي با عنوان "قواعد وابستگي: استخراج دانش با تحليل سبد خريد"، دريافتند كه وابستگي مشخصي بين محصولات خريداري شده در زمان خريد توسط مشتريان وجود دارد كه داده‌كاوي مي‌تواند روابط ميان محصولات را تعيين و الگوي خريد مشتريان را پيشبيني كند و در مكان‌يابي محصولات در قفسه‌ها به كار رود. آني‌ام‌سي و كوماردي (۲۰۱۲)، در پژوهش خود با عنوان "تحليل سبد خريد براساس استخراج مجموعه اقلام تكرارشونده در فروشگاه‌هاي بزرگ"، به اين نتيجه رسيدند كه به كارگيري تكنيك‌هاي داده‌كاوي در كشف اقلام مرتبط باعث افزايش رضايت مشتريان و به دنبال آن افزايش فروش و سود فروشگاه مي‌شود. مگيايز و همكاران (۲۰۱۳)، در تحقيقي با عنوان "مدل‌سازي بخشي از مشتريان متغير: ارزيابي اولين گروه محصولاتي كه با ترتيب خريداري مي‌شوند"، به اهميت نگهداري مشتريان به ويژه در بخش خرده‌فروشي تأكيد نموده و با استفاده از تكنيك‌هاي داده‌كاوي، مدل‌هاي پيشگويانه را براي تغييرات در مشتريان ارائه كرده‌اند كه شركت‌ها مي‌توانند با استفاده از آن هدف‌گيري آينده خود را جهت حفظ موضع رقابتي بازار تعريف كنند. در تحقيقي كه توسط گوپتا و گوپتا (۲۰۱۴)، در كشور هندوستان، تحت عنوان "تحليل رفتار مشتريان با استفاده از تكنيك‌هاي داده‌كاوي: بهينه‌سازي ارتباط با مشتري"، انجام گرفت به كاربرد داده‌كاوي جهت درك و پيش‌بيني مسائل اصلي رفتار مشتريان با هدف بهبود تصميم‌گيري و ارتباط با مشتري در كسب‌وكارهاي مشتري‌مدار تأكيد شده‌است. باتچر و همكاران (۲۰۰۹)، مطالعه‌اي تحت عنوان "كاوش تغييرات بخش‌هاي مشتريان در بازارهاي پويا" انجام دادند؛ در اين تحقيق سيستمي براي بخش‌بندي مشتريان در بازارهاي پوياي امروزي ارائه شده‌است. روش ارائه شده در اين پژوهش مبتني بر كشف مجموعه اقلام مكرر و تحليل تغييرات آنها در طول زمان است و امكان يافتن بخش‌هاي دلخواه و تجزيه و تحليل توسعه زماني آن‌ها را فراهم مي‌كند. انگاي و همكاران (2009)، در پژوهشي كم نظير با عنوان "كاربرد تكنيك‌هاي داده‌كاوي در مديريت ارتباط با مشتري"، مطالعات صورت گرفته در سال‌هاي 2006-2000 پيرامون كاربرد تكنيك‌هاي داده‌كاوي در مديريت ارتباط با مشتري را با هدف ارائه خلاصه‌اي در اين زمينه و روش‌هايي داده‌كاوي كه اغلب مورد استفاده قرار مي‌گيرد، جمع‌آوري، دسته‌بندي و مطالعه كرده‌اند. در مطالعه‌اي ديگر با عنوان "كاربرد داده‌كاوي در بهبود مديريت ارتباط با مشتري" كه توسط تارخ و شريفيان (۱۳۸۹) انجام گرفت، با استفاده از تكنيك‌هاي داده‌كاوي مشتريان بانك ملت شناسايي و بخش‌بندي شده و يك سيستم خبره توليد كه مديريت را در تصميم‌گيري و اخذ استراتژي‌هاي كارا براي برخورد با مشتريان پشتيباني مي‌كند، ارائه دادند. تقوي‌فرد و خواجوند (۱۳۹۲)، در مطالعه‌اي تحت عنوان "بخش‌بندي خوشه‌اي مشتريان بانك"، با استفاده از داده‌كاوي و از طريق محاسبه ارزش طول عمر مشتريان به كشف الگوهاي پنهان ميان داده‌هاي استفاده از خدمات بانكداري و بخش‌هاي مختلف مشتريان پرداختند. نتايج اين تحقيق راهكارهايي را جهت تدوين استراتژي‌هاي بازاريابي و پيشنهاد محصول و خدمات براي هريك از گروه‌هاي مشتريان ارائه مي‌دهد. فشاركي و همكاران (۱۳۸۹)، در مطالعه‌اي تحت عنوان "استخراج دانش از پايگاه داده نرم‌افزارهاي مديريت اطلاعات و مستندات با تكنيك‌هاي داد‌هكاوي"، دريافتند با استفاده از تكنيك‌هاي داده‌كاوي مي‌توان از بانك‌هاي اطلاعاتي، دانش مورد نياز را به دست آورد. اين پژوهش چارچوب استاندارد داده‌كاوي را براي پايگاه داده معرفي كرده‌است. ديوسالار و بهبود (۱۳۸۵)، در مقاله خود با عنوان "مديريت بر اطلاعات مشتري در مديريت زنجيره تأمين با استفاده از رويكرد داده‌كاوي"، چگونگي استفاده از داده‌كاوي به عنوان يك تكنيك كارا براي استخراج دانش در يك چارچوب مديريت دانش مشتري را نشان مي‌دهند. اين مطالعه داده‌كاوي را ابزار و روشي جهت كشف دانش و درك بهتر مشتري معرفي مي‌كند.

* + 1. تعریف موضوع تحقیق

اغلب الگوریتم‌های یادگیری ماشین در داده‌کاوی با داده‌های عددی کار می‌کنند و در پیاده‌سازی و نحوه کار آن‌ها گرایش به ریاضیات محض وجود دارد. اما، کاوش قواعد وابستگی (association rule mining) که از آن با عنوان کاوش قواعد انجمنی نیز یاد می‌شود، برای داده‌های دسته‌ای مناسب و محاسبات آن نسبت به بسیاری از دیگر الگوریتم‌ها ساده‌تر است. این روش، یکی از راهکارهای مبتنی بر قواعد (rules)، برای کشف روابط جالب بین متغیرها در پایگاه داده‌های 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 می‌آید دید.

* I = {i1, i2, …} مجموعه‌ای از ویژگی‌های 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،­­ t1 مقدم و t2 نتیجه (موخر) محسوب می‌شود.

* t1 ⇒ t2 (در اینجا، ti به طور کلی یک مورد مجزا یا مجموعه‌ای از اقلام است.)
  + 1. هدف یا هدف‌های کلی و آرمانی تحقیق

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 بالا رفتن وفاداری مشتری می‌شود. با معرفی مناسب محصولات، فضای رشد محصولات جدید پر رونق‌تر می‌شود و این مساله به بازاریابی تولیدی‌های نوپا یا محصولات کمتر شناخته شده در کسری از جامعه کمک می‌کند.

* + 1. روش انجام پروژه

در حالت کلی ابتدا به دنبال داده‌های مناسب می‌گردیم که بتوان با آن نزدیکترین شبیه‌سازی را با داده‌های واقعی کرد. سپس به پیاده‌سازی الگوریتم آن با استفاده از الگوریتم‌های داده‌کاوی می‌پردازیم. بعد از آماده شدن الگوریتم به درست کردن بخش سرور اپلیکیشن می‌پردازیم و در ادامه آن به درست کردم یک صفحه که بتوان این کارها را در آن نشان داد.

برای داده‌ها از داده‌های دانشگاه UCI استفاده می‌کنیم[[1]](#footnote-1). این داده‌ها شامل تراکنش‌های سفارش‌های یک مغازه است. سپس برای شروع از الگوریتمی به نام اپریوری (Apriori) استفاده می‌شود که بتوانیم قواعد انجمنی را استخراج کنیم و همچنین در آخر از بقیه الگوریتم‌ها نیز استفاده می‌کنیم که بتوانم بهترین پاسخ را انتخاب کنم.

در ادامه به سازگار کردن این کدها با بخش سرور نرم‌افزار یا همان بخش back-end با استفاده از Flask Restful می‌پردازیم. در آخر نیز به طراحی صفحاتی با استفاده از ابزار Vue-js پرداخته می‌شود که اینکار توسط خانم مینو احمدی انجام می‌گیرد.

* + 1. نوآوری، اهمیت و ارزش تحقیق

در این تحقیق سعی شده تا با پیوند داده‌کاوی و بازاریابی و پیاده‌سازی آن‌ها در بستر وب اپلیکیشن نتایج بسیار ارزنده بدست آمده را در عمل در دسترس شرکت‌های بزرگ قرار دهد. عملی کردن دسترسی آسان به این نتایج نوآوری‌ای بوده که تا به حال انجام نگرفته است.

همچنین در تحقیق‌های پیشین صورت گرفته برای استفاده از قوانین انجمنی و پیدا کردن نتایج مناسب فقط از یک الگوریتم استفاده شده است ولی ما در این تحقیق با به کارگیری و بررسی معایب و مزایای چندین الگوریتم سعی کرده‌ایم تا بهترین نتایج را بدست بیاوریم.

ارزش این تحقیق در استفاده عملی و ساده و آسان شرکت‌ها بر بستر وب و بر پایه الگوریتم‌های مختلف داده‌کاوی است که منجر به بهبود بازاریابی و افزایش سود و بهره‌وری شرکت‌ها می‌شود.

* + 1. خلاصه فصل‌ها

در فصل بعدی مروری بر کارهای انجام شده خواهیم داشت و همچنین مروری بر تعاریف الگوریتم‌ها و معیارهای استفاده شده در پروژه می‌کنیم. در فصل سوم به نحوه پیاده‌سازی پروژه می‌پردازیم که نتایج آن در فصل چهارم نشان داده خواهد شد. در فصل پنجم به بحث و نتیجه‌گیری حول این پروژه می‌پردازیم.

1. مروري بر مطالعات انجام شده
   1. مقدمه

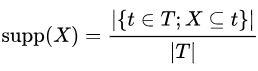
فرآيند كشف قوانين وابستگي، يكي از رويكردهاي مهم در علم نوين داده كاوي براي يافتم قواعد و الگوها در پايگاه داده است كه بسيار مورد توجه تحقيقگران قرار گرفته است . يكي از كاربردي ترين حالات تحليل وابستگي ها "تجزيه و تحليل سبد بازار" مي باشد كه در آن هدف يافتن كالاهايي است كه معمولا به طور همزمان خريداري مي شوند. پيشرفت فناوري، فروشگاه هاي خرده فروش را قادر ساخته است تا حجم زيادي از داده هاي خريد مشتريان (كه از آن به عنوان سبد خريد ياد مي شود) را جمع آوري و ذخيره نمايد. هر مشتري خريد مجزايي را در مقادير مختلف و زمان هاي متفاوت انجام مي دهد و داده هاي موجود در سبد بازار، نشان دهنده خريد مشتري در يك زمان خاص است. با تجزيه و تحليل سبد بازار خرده فروشان، رفتار خريد مشتريان را پيش بيني كنند. اين كار به آنها كمك مي كند تا بتوانند كالاهاي خود را بهتر ساماندهي كرده و چيدمان بهتري از محصولات خود داشته باشند و از اين طريق سودآوري خود را افزايش دهند. هدف اصلي داده كاوي در پيدا كردن وابستگي، يافتن قانون هاي محكم و قابل توجه است. قوانين وابستگي اولين بار در سال 1993 براي مقادير صفر و يك توسط اگراوال و همکاران مطرح شد.

* 1. تعاريف، اصول و مباني نظري

در داده‌کاوی و به ویژه کاوش قواعد وابستگی، به منظور انتخاب قواعد جالب از میان مجموعه‌ای از قواعد ممکن، محدودیت‌های گوناگونی (به عنوان آستانه) بر سنجه‌های مختلف مرتبط با اهمیت و جالبی (interestingness)، اعمال می‌شود. شناخته شده‌ترین محدودیت‌ها در کاوش قواعد وابستگی، آستانه کمینه برای پشتیبانی (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 تراکنش‌های t در مجموعه داده‌ای که شامل مجموعه اقلام X است تعریف می‌شود. درواقع پشتیبان X معادل است با کسر تعداد دفعاتی که قلم جنس X در تراکنش‌ها (سفارشات) حاضر شده است.



### 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.

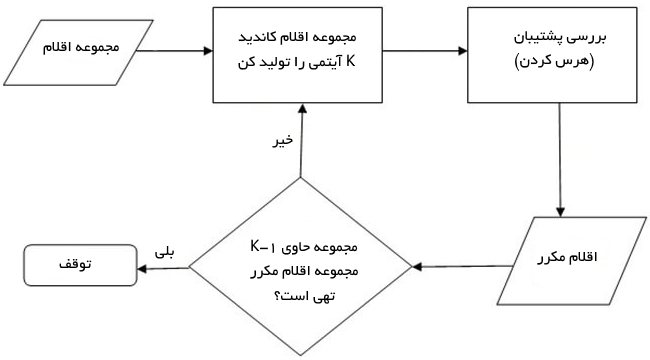
اطمینان در داده‌کاوی

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

* 1. مروري بر ادبيات موضوع

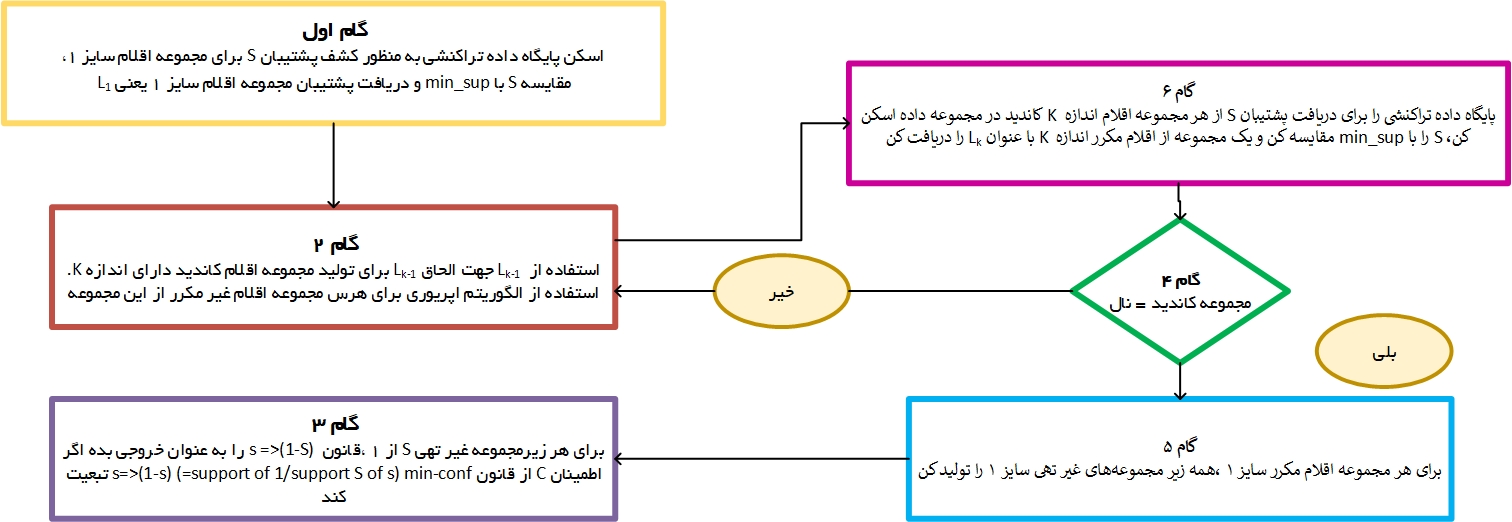
### الگوریتم Apriori

الگوریتم اپریوری (Apriori) بر این اصل بنا شده که اگر یک مجموعه اقلام (itemset) مکرر است، پس همه زیرمجموعه‌های آن نیز مکرر هستند. این بدین معنا است که اگر {0,1} مکرر باشد، پس {0} و {1} نیز مکرر هستند. بالعکس این قاعده نیز صادق است، یعنی اگر یک مجموعه اقلام مکرر نباشد، زیرمجموعه‌های 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 داده‌های عددی و اسمی (دسته‌ای) است.



شکل 1- الگوریتم اپریوری۱

برای کسب درک بهتر از الگوریتم می‌توان برخی کاربردهای آن مانند «تحلیل سبد خرید» را مورد بررسی قرار داد. در این کاربرد، داده‌کاو به دنبال کشف آن است که کدام اقلام با یکدیگر (در یک سبد خرید) خریداری شده‌اند. 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.



شکل 2-الگوریتم اپریوری ۲

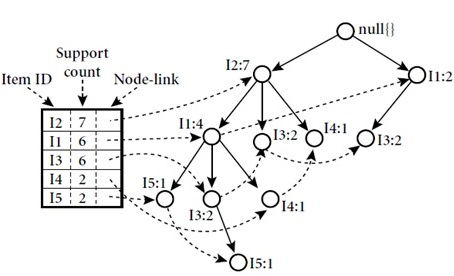
این روش، ممکن است به طور پیوسته با دیگر الگوریتم‌ها استفاده شود تا به طور موثری داده‌ها را مرتب‌سازی و با یکدیگر مقایسه کند. اصل اپریوری می‌تواند تعداد اقلامی که نیاز به بررسی آن‌ها است را کاهش دهد. این روش بیان می‌کند که اگر یک مجموعه اقلام فاقد تکرار است، پس همه زیرمجموعه‌های آن نیز نادر هستند. این 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.

### الگوریتم FP-Growth

تاکنون الگوریتم‌های بسیاری در جهت کاوش قواعد انجمنی ارائه شده که FP-growth یکی از موفق‌ترین آن‌هاست. این الگوریتم داده‌های موجود در پایگاه داده را به شکلی فشرده در درختی موسوم به FP-tree ذخیره و سپس با ساخت بازگشتی FP-tree قلم‌داده‌های مکرر را کشف می‌کند. این الگوریتم روشی است که مجموعه اقلام پرتکرار را بدون تولید مجموعه اقلام کاندید به­ دست می­آورد که از یک استراتژی تقسیم وحل استفاده می­کند. این روش پایگاه ­داده را به مجموعه‌­ای از پایگاه­داده­ها که هرکدام یک قلم پرتکرار دارند، تقسیم می­کند و هر پایگاه ­داده را جداگانه کاوش می­کند.

در اولین اسکن پایگاه ­داده همانند الگوریتم اپریوری مجموعه آیتم­های یک­ عضوی و پشتیبانی آنها مشخص می­شود. مجموعه اقلام پرتکرار به ترتیب نزولی پشتیبانی­شان مرتب می­شوند.

سپس یک درخت به این صورت ساخته می­شود که: اوّل ریشه درخت با برچسبnull ساخته می­شود. بعد از آن پایگاه ­داده برای بار دوم اسکن می­شود، اقلام هر تراکنش به ­ترتیب L پردازش می­شوند و یک شاخه برای هر تراکنش ایجاد می­شود. به­ منظور تسهیل پیمایش درخت، یک جدول ساخته می­شود که هر قلم درآن به محل­ خودش در درخت ­اشاره می­کند. درخت پس از اسکن همه تراکنش‌ها کامل می­شود.



شکل 3-الگوریتم FP-Growth

* 1. نتيجه‌گيري

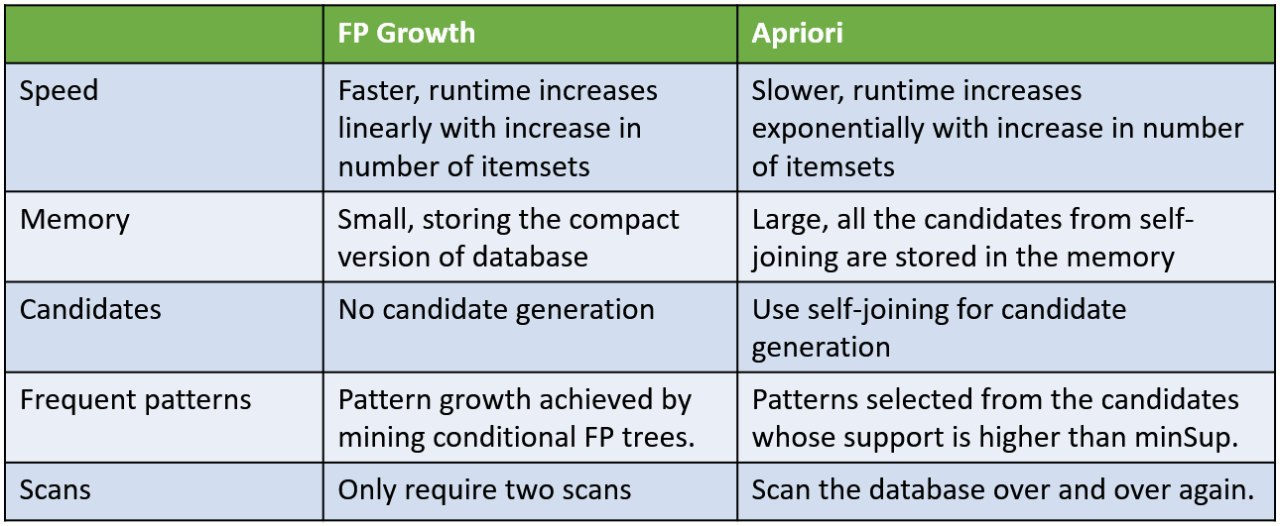
### محدودیت‌های الگوریتم Apriori

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 قرار گیرد.

### Apriori vs FP-Growth

می‌توانیم تفاوت این دو الگوریتم را به صورت خلاصه شده در جدول زیر ببینیم:



شکل 4- تفاوت الگوریتم اپریوری و FP-Growth

همانطور که می‌دانیم به طور کلی الگوریتم FP-Growth از الگوریتم Apriori بهتر است و الگوریتم FP-Growth در واقع بهبود یافته الگوریتم اپریوری است.

1. روش تحقيق
   1. مقدمه

برای راه اندازی این وب اپلیکیشن ما کارها را به چند بخش تقسیم کردیم. ابتدا به آماده‌سازی الگوریتم مورد نیاز پرداختیم. سپس با استفاده از این الگوریتم یک سرویس آماده می‌کنیم که بتوانیم از آن امکاناتی که آماده کردیم استفاده کنیم. سپس برای این سرویس یک رابط کاربری گرافیکی می‌سازیم. داده‌های اینپروژه بر اساس داده‌های واقعی بدست آمده توسط دانشگاه UCI است.

* 1. الگوکاوی فراوانی
     1. علت انتخاب روش

همانطور که در فصل 1 گفته شد، اغلب الگوریتم‌های یادگیری ماشین در داده‌کاوی با داده‌های عددی کار می‌کنند و در پیاده‌سازی و نحوه کار آن‌ها گرایش به ریاضیات محض وجود دارد. اما، کاوش قواعد انجمنی برای داده‌های دسته‌ای مناسب‌تر و محاسبات آن نسبت به بسیاری از دیگر الگوریتم‌ها ساده‌تر است. این روش، یکی از راهکارهای مبتنی بر قواعد (rules)، برای کشف روابط بین متغیرها در پایگاه داده‌های بزرگ محسوب می‌شود.

* + 1. تشريح كامل روش تحقيق

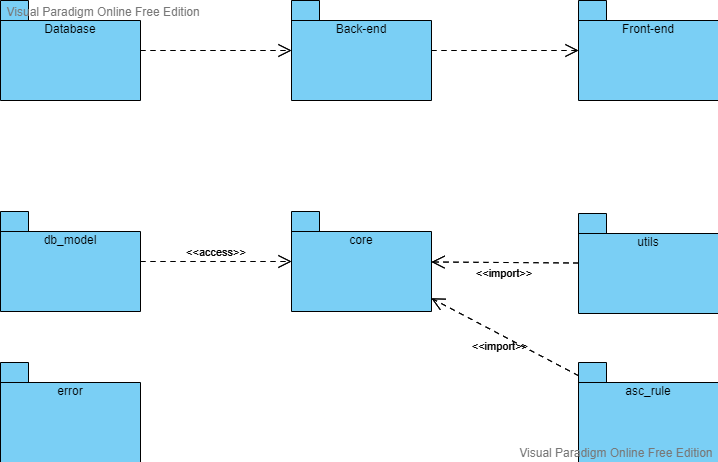
در اولین قدم ما نیاز داشتیم که الگوریتم مورد نیاز برای دو قابلیت گفته شده را آماده‌سازی کنیم. در اولین قابلیت می‌خواهیم به کاربر بر اساس سبد خرید انتخابی خود، کالاهایی را پیشنهاد دهیم که بقیه کاربران در خریدهای مشابه با این کاربر آن‌ها را خریداری کردند. دومین قابلیت این امکان را به کاربر می‌دهد که بر اساس کالاهایی که انتخاب کرده، مشتری مناسب برای فروش آن‌ها را پیدا کند.

ابتدا داده‌ها را بر اساس کشور مورد نظرمون فیلتر می‌کنیم. سپس اسامی کالاها را تمیز می‌کنیم که بتوانیم در ادامه مسیر با آسایش خاطر از آن‌ها استفاده کنیم مانند حذف کردن چند فاصله پشت هم یا حذف فاصله‌های ابتدا و انتهای کلمات. حال نیاز داریم قوانین انجمنی حاکم بر این کالاها را بدست بیاوریم که از روی آن‌ها قابلیت‌های گفته شده را درست کنیم. در ابتدا نیاز داریم که داده‌های کالاها را به صورت یک ماتریس دربیاوریم که هر ستون آن نشان دهنده یک کالا و هر سطر آن نشان‌دهنده یک شخص است و هر درایه آن نشان می‌دهد که آیا کالای ستون j ام را شخص سطر i ام خریداری کرده است یا خیر. یک حالت دیگری که می‌توانستیم در نظر بگیریم وجود هر کالا در هر فاکتور خرید بود ولی در این حالت بدلیل اینکه تعداد کالاها در هر فاکتور کم است، با اینکار نتیجه مطلوبی گرفته نمی‌شود. پس از ساخت این ماتریس ما فراوانی هر کالا را بدست می‌آوریم که اینکار را می‌توان با الگوریتم‌های مختلفی انجام داد که در این پروژه از دو الگوریتم apriori و fp-growth استفاده شده است. پس از پیدا کردن فراوانی کالاها به پیدا کردن قوانین انجمنی می‌پردازیم. وقتی قوانین را پیدا کردیم، برای اینکه قوانین بهتر را شناسایی کنیم و از آن‌ها استفاده کنیم، روی معیارهایی که در فصول قبل معرفی و تعریف شدن شروط حداقلی می‌گذاریم که به این معنی است که اگر یک قانون از یک حدی ضعیف‌تر باشد آن قانون حذف می‌شود. در انتهای این بخش ما قوانین خوبی را خواهیم داشت که در ادامه از آن‌ها استفاده می‌کنیم.

برای پیاده‌سازی قابلیت اول فرض می‌کنیم که یک لیستی از کالاها داریم که نشان‌دهنده سبد خرید کاربر است. حالا می‌خواهیم بر اساس این سبد و قوانینی که بدست آوردیم یک سری کالا به کاربر پیشنهاد شود. برای اینکار لیستی از مجموعه همه حالات قرارگیری کالاها با داشتن یک یا دو کالا بدست می‌آوریم که هر عضو این لیست را با قوانینی که در مرحله قبل بدست آورده بودیم مقایسه می‌کنیم. اگر مطابقت داشتند و خروجی آن قانون در سبد نبود، آن کالا به عنوان کالای پیشنهادی خروجی داده می‌شود.

برای قابلیت دوم، لیست ورودی کالاها را به جای مقایسه با ورودی یک قانون، با خروجی قوانین مقایسه می‌کنیم و افرادی که کالاهای ورودی قوانین را خریداری کرده‌اند نمایش می‌دهیم. اینکار به این معنی است که این اشخاص کالاهایی را خریده‌اند که مشابه با کالاهای کاربر بوده ولی آن کالاها را نخریده‌اند و به احتمال خوبی به این کالاهای جدید نیاز دارند پس گزینه‌های خوبی برای فروش این کالاها هستند.

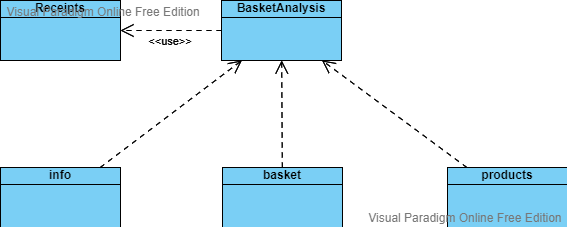
حالا برای استفاده از این قابلیت‌ها نیاز داشتیم که یک سرویس نوشته شود. برای پیاده‌سازی این سرویس از Flask-restful استفاده شده است. برای ساخت آن ابتدا نیاز داشتیم که یک طراحی برای بخش back-end آن داشته باشیم. این طراحی به صورت زیر است:



شکل 5– نمودار پکیج‌ها

در بخش بالایی این دیاگرام نمای کلی این نرم‌افزار است که نشان می‌دهد این نرم‌افزار از سه بخش اصلی دیتابیس یا پایگاه داده، بخش سرویس back-end و بخش front-end تشکیل شده است. روابط دو بخش پایگاه داده و back-end به صورت بخش پایینی این عکس است. پس از طراحی و پیاده‌سازی این بخش‌ها و متصل کردن این سرویس به دیتابیس (db\_model)، نیاز داشتیم خطاهایی که در الگوریتم وجود دارد در سیستم نمایش داده شود. اینکار توسط فایل error انجام گرفت که خطاهای مرتبط در آن تعریف می‌شود و در جای مناسب نمایش داده می‌شود. مثالی از این خطاها می‌تواند خروجی نداشتن یکی از قابلیت‌ها باشد. از آنجایی که ما قوانینی که معیارهای خوبی را دارا هستند نگه داشتیم، لزوما با انتخاب هر مجموعه‌ای از کالاها خروجی نداریم. پس نیاز به پیام مناسب از طرف سرور بود.

برای داشتن دید دقیقتر به طراحی شی‌گرای این نرم‌افزار در سمت سرور می‌توانیم از دیاگرام زیر استفاده کنیم که نشان‌دهنده طراحی آن است.



شکل 6-نمودار طراحی شی‌گرا

پس از پیاده سازی بخش back-end ما قسمت front-end را برای نمایش انچه در بک اند انجام شده و پیاده سازی ان به طوری که استفاده از ان مانند یک اپلیکشن تحت وب ساده باشد پیاده سازی میکنیم .

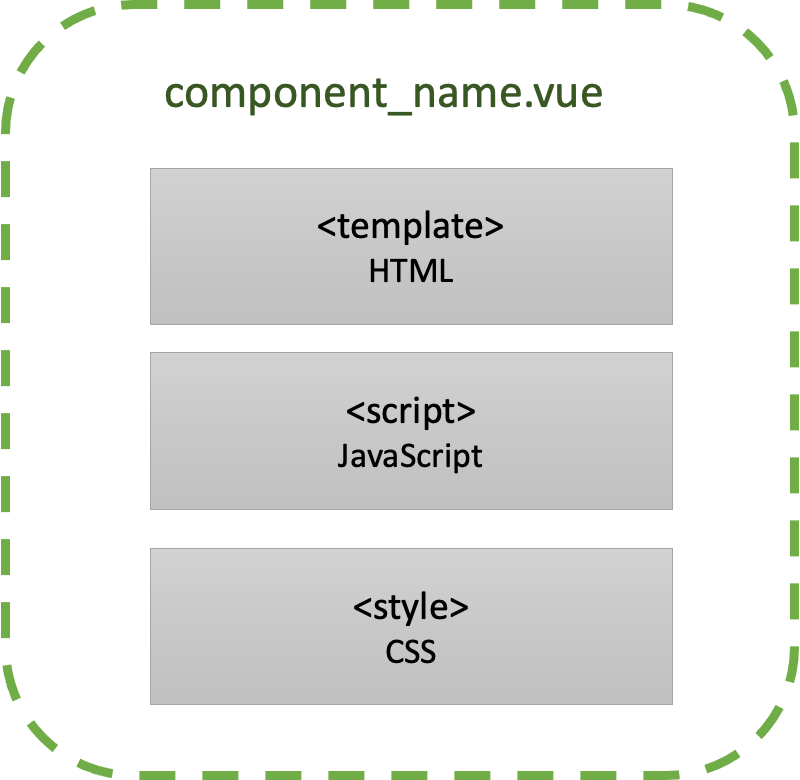
در این پروژه قسمت front-end را به کمک فریم ورک معروف جاوا اسکریپت به نام Vue.js پیاده سازی کردیم چرا که این فریم ورک دارای قابلیت های ویژه و مناسبی می باشد.

این فریم ورک component base است همه ی بخش ها حالت کامپوننت طراحی می‌شوند ساختار هر component در ویو سه بخش دارد :

1- تمپلیت :بخش یو ای یا همان کد های اچ تی ام ال است.

2- اسکریپت : کد های جاوا اسکریپت است.

3-استایل : که در آن کد های cssو استایل دهی هستند.



شکل 7- بخش های مختلف ویو

در پیاده سازی این بخش ابتدا اجزای مختلف انچه باید در سایت نمایش داده شود (بخش های مختلف سایت مانند footer, header و.. ) را طراحی و پیاده سازی کردیم و سپس به کمک کتابخانه Router این قسمت ها را به هم لینک کرده و مربوط می سازیم پس از لینک کردن و مرتبط ساختن تمام بخش های front به یک دیگر به کمک کتابخانه axios انچه در front پیاده سازی کردیم را به back وصل کرده و به هم مرتبط می سازیم.

در شکل زیر میتوان رابطه ی back-end و front-end را به خوبی مشاهده کرد .

ذتذ

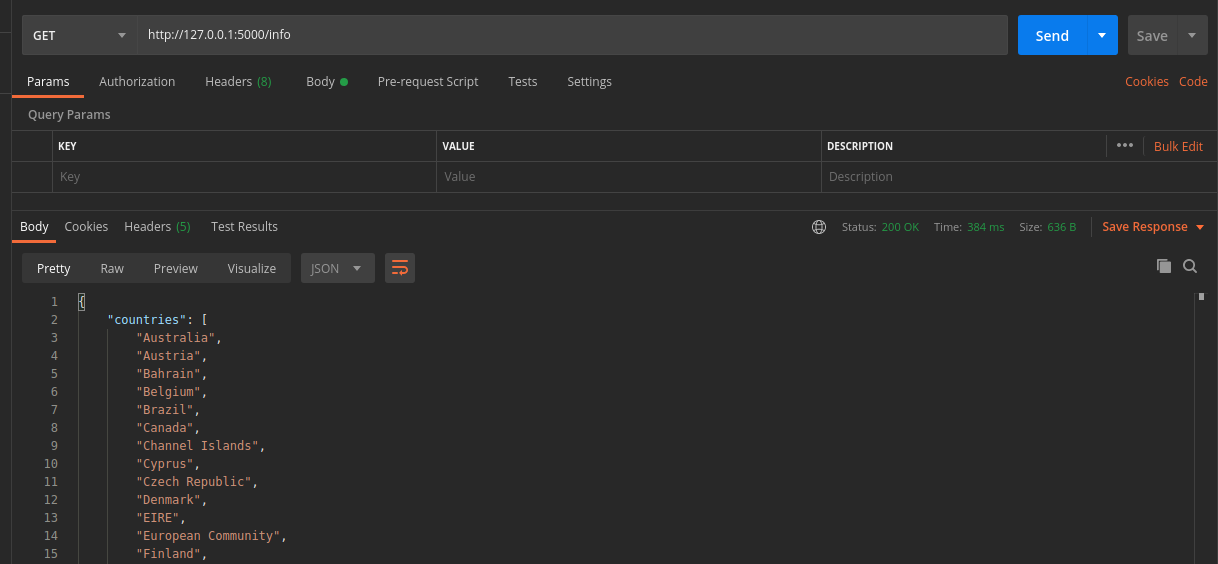

شکل 8-نحوه ارتباط front-end و back-end

1. نتايج
   1. مقدمه

در این پروژه ما سعی کردیم تا علاوه بر استفاده از الگوریتم‌های متفاوت، آن‌ها را در بستر یک وب اپلیکشین ساده پیاده‌سازی کنیم تا به راحتی با اتصال آن به پایگاه‌داده‌ها بتوانیم پیشنهادات محصول و مشتریان را استخراج کنیم که در ادامه این نتایج آورده شده است.

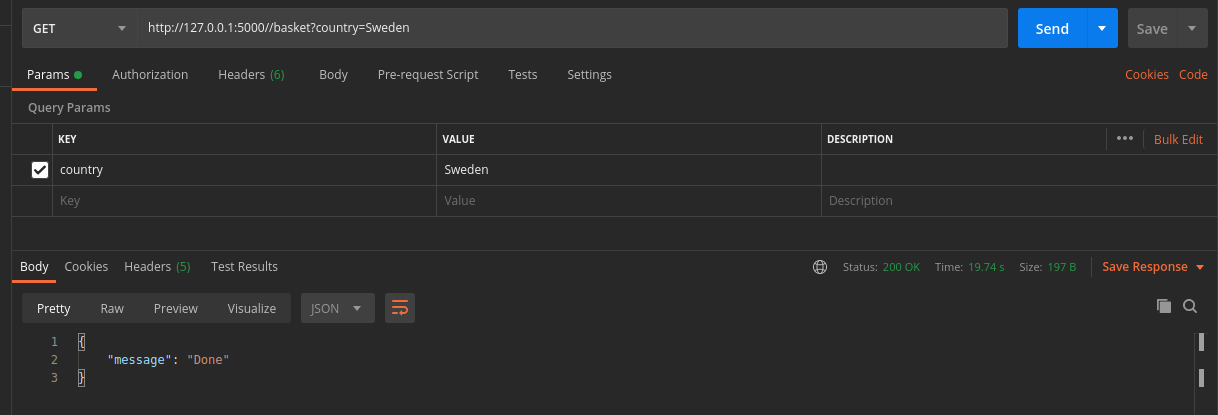
* 1. محتوا

برای نشان دادن خروجی سرویس پیاده شده بر روی الگوریتم‌ها، از نرم‌افزار Postman استفاده می‌کنیم. متدهای پیاده شده برای نشان دادن داده‌ها و فراخوانی و اجرای دو قابلیت است. در عکس زیر می‌توانیم متد چاپ همه کشورها را ببینیم.



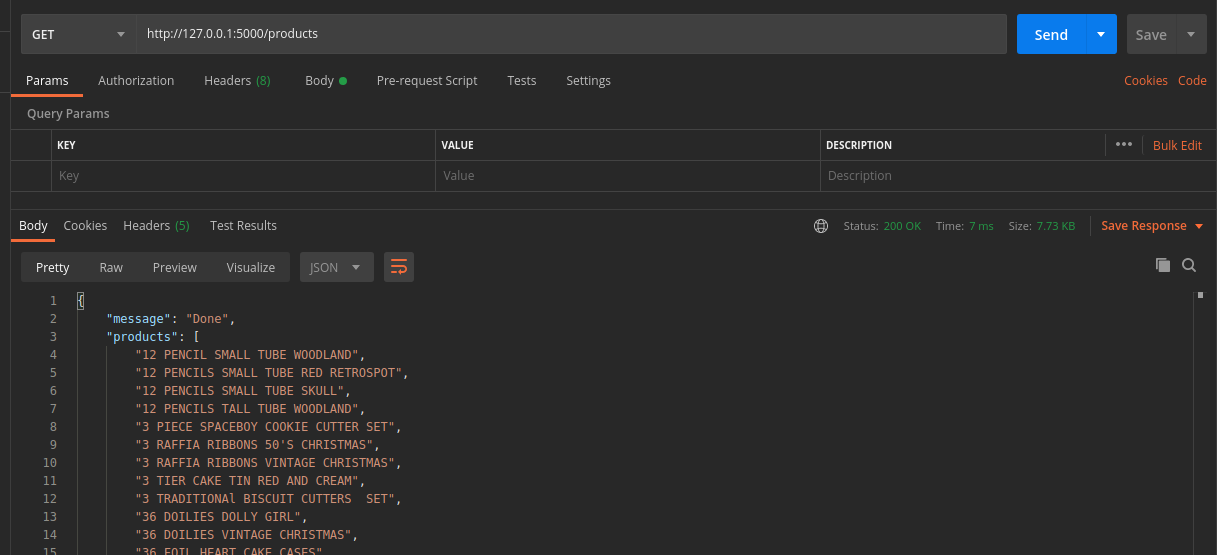
شکل 9 - لیست کشورها

پس از گرفتن کشورها، نیاز داریم کشوری را که در آن می‌خواهیم سفارش‌ها را بررسی کنیم انتخاب کنیم. برای انتخاب کشور از متد زیر استفاده می‌کنیم که در مثال پایین کشور سوییس انتخاب شده است. در این متد علاوه بر انتخاب کشور، در سمت سرور قوانین انجمنی نیز ساخته می‌شود که در مراحل بعدی مورد استفاده قرار می‌گیرند.



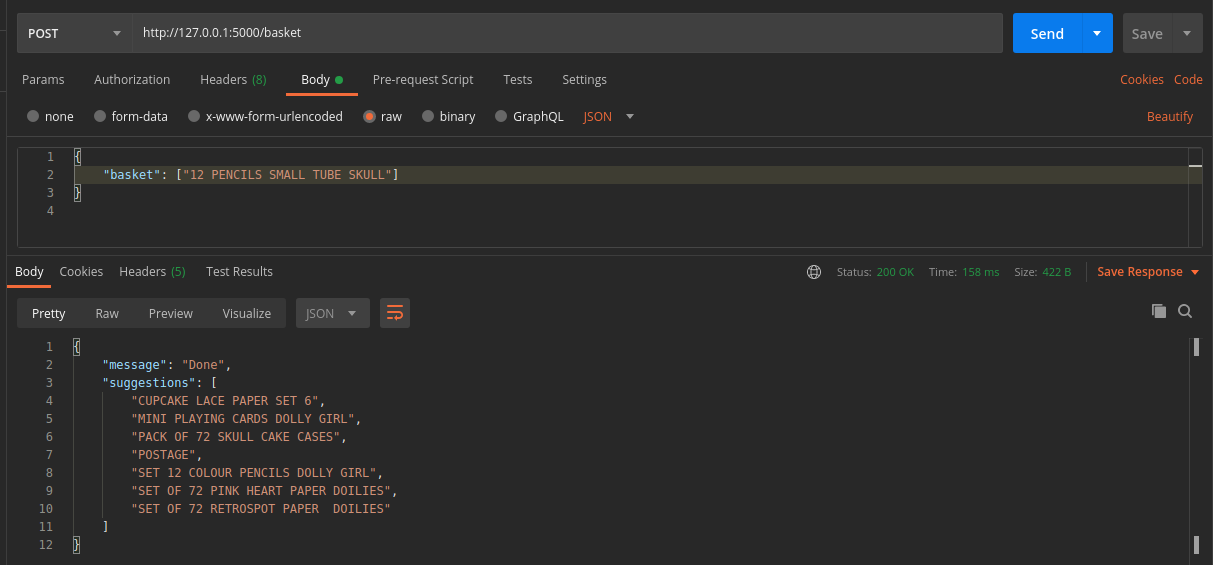
شکل 10- انتخاب کشور مورد نظر

بعد از اینکه قوانین مربوط به آن کشور ساخته شد، نیاز داریم تا کالاهای موجود در فروشگاه را مشاهده کنیم که لیستی از کالاهای مورد نظر خودمان را از بین آن‌ها انتخاب کنیم. با استفاده از متد زیر می‌توانیم لیست کالاهای فروشگاه را مشاهده کنیم.



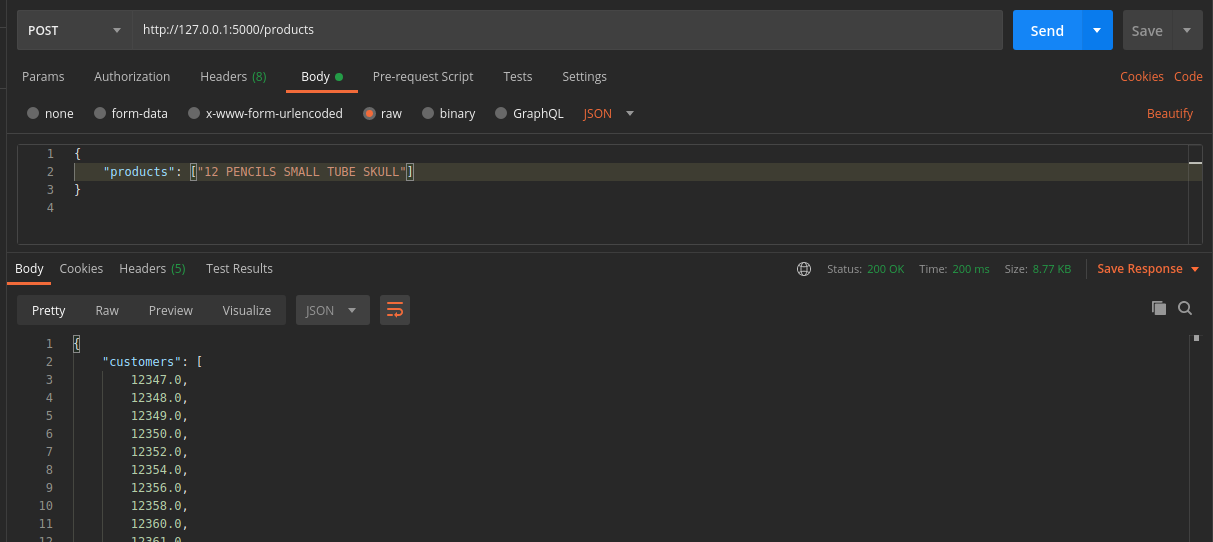
شکل 11– لیست محصولات

حال می‌توانیم چندین کالا را انتخاب کنیم و با استفاده از متد زیر از قابلیت اول استفاده کنیم. قابلیت اول به ما لیستی از کالاها را بر اساس سبد فعلیمان پیشنهاد می‌دهد. در اینجا سبدمان شامل کالای 12 PENCILS SMALL TUBE SKULL است و در پایین کالایهای پیشنهادی را می‌توانیم ببینیم.



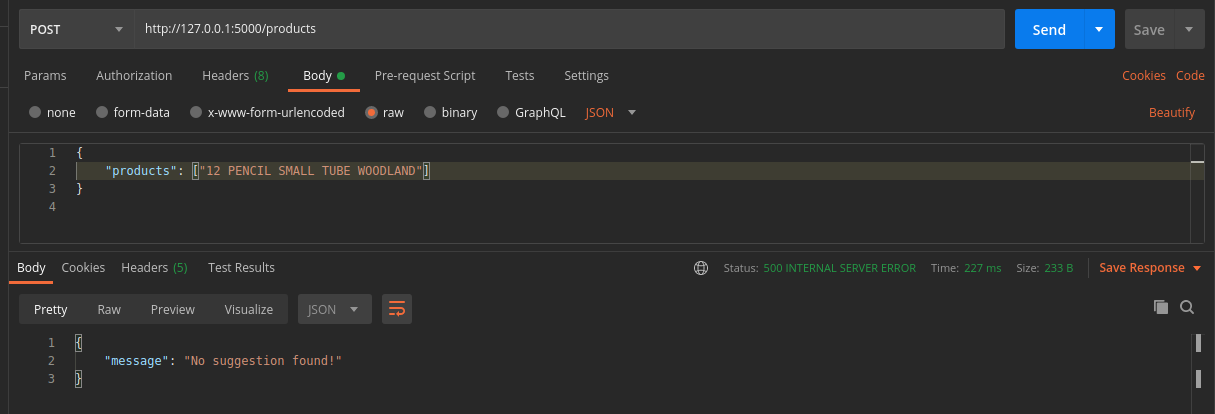
شکل 12- پیشنهاد محصول

اگر بخواهیم از قابلیت دوم استفاده کنیم می‌توانیم از متد زیر استفاده کنیم. این قابلیت این امکان را به ما می‌دهد که با انتخاب چند کالای مورد نظر، بهترین افرادی که ممکن است آن‌ها را خریداری کنند را پیدا کنیم. در مثال زیر نیز از کالای مثال قبل استفاده کردیم.



شکل 13– پیشنهاد مشتری

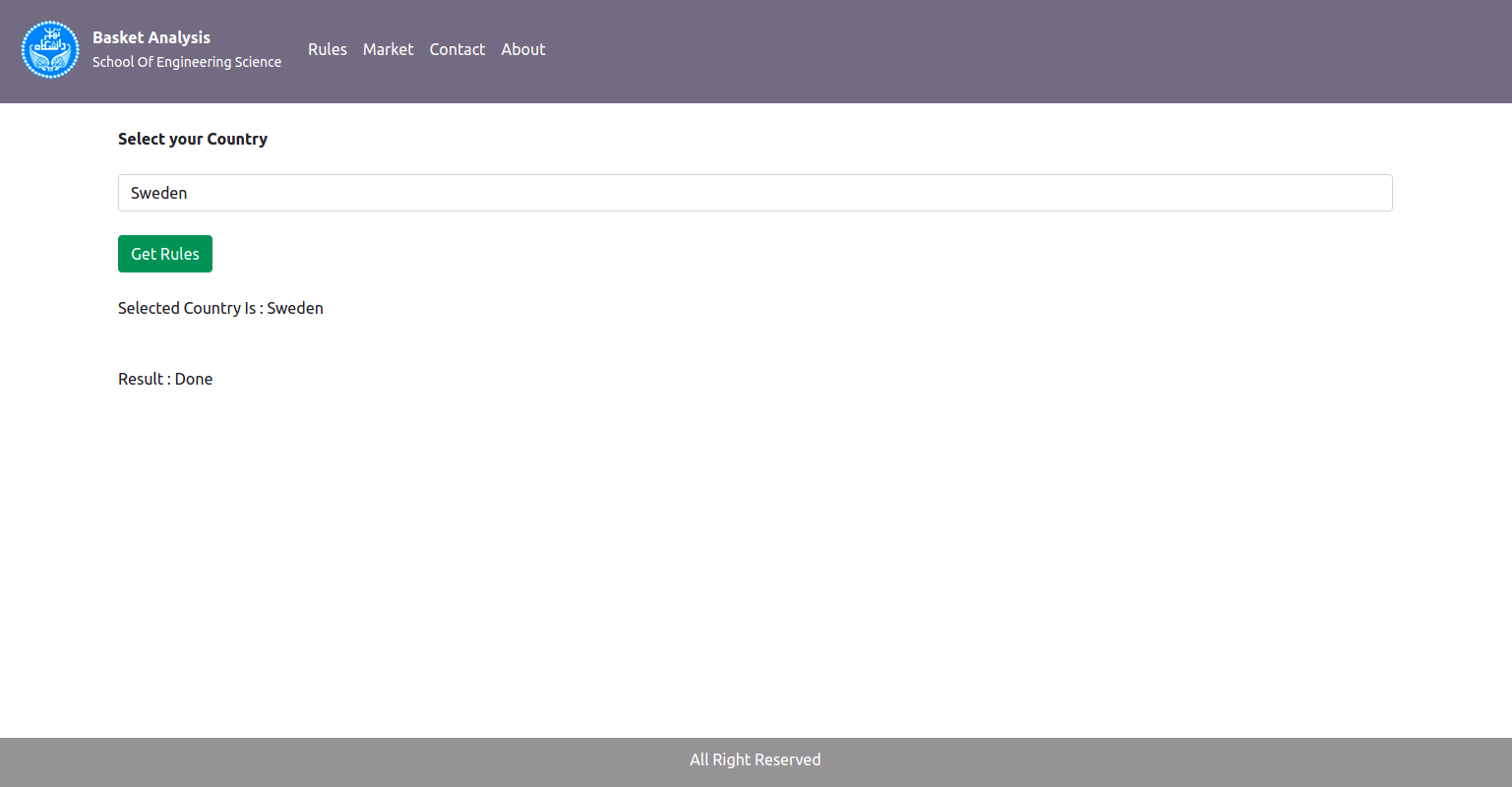
در این سرویس اگر ما پیشنهادی را بخواهیم ولی قانونی برای آن مجموعه از کالاها وجود نداشته باشد، خطایی بازگشردانده می‌شود مبنی بر این که پیشنهادی یافت نشد. این خطا را در تصویر زیر می‌توانیم ببینیم.



شکل 14– نمایش خطا

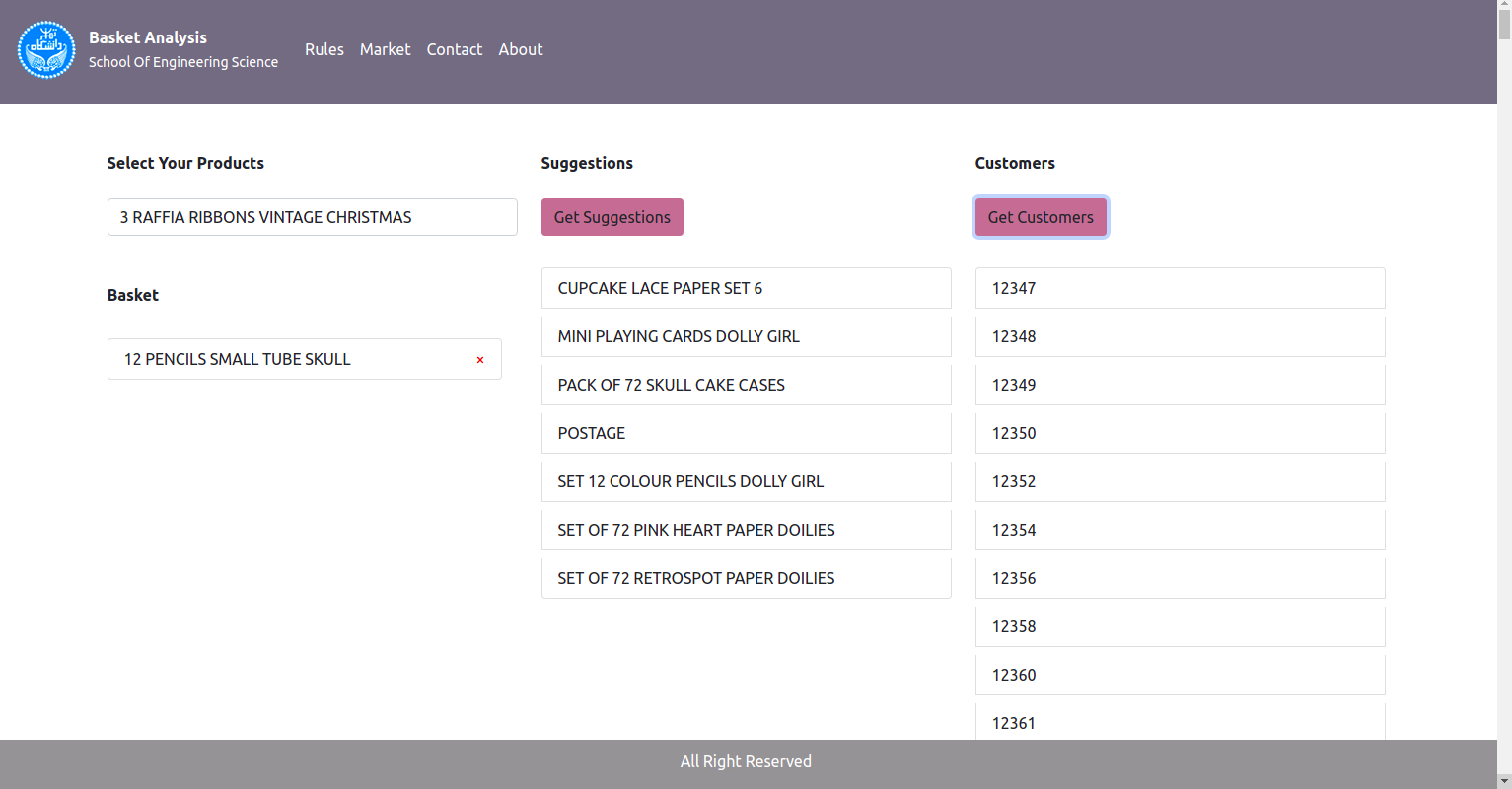
این نتایج فقط برای سرویس back-end بودند. اگر بخواهیم این نتایج را به صورت گرافیکی نیز نشان دهیم می‌بایست یک front-end برای آن طراحی و اجرا کنیم. خروجی این سرویس گرافیکی را در ادامه نشان می‌دهیم.

تصویر زیر نشان دهنده بخش انتخاب کشور است که بعد از انتخاب گزینه Get Rules قوانین در سرور ساخته می‌شود. اگر این امر موفقیت‌آمیز باشد در تصویر کلمه Done را مشاهده می‌کنیم.



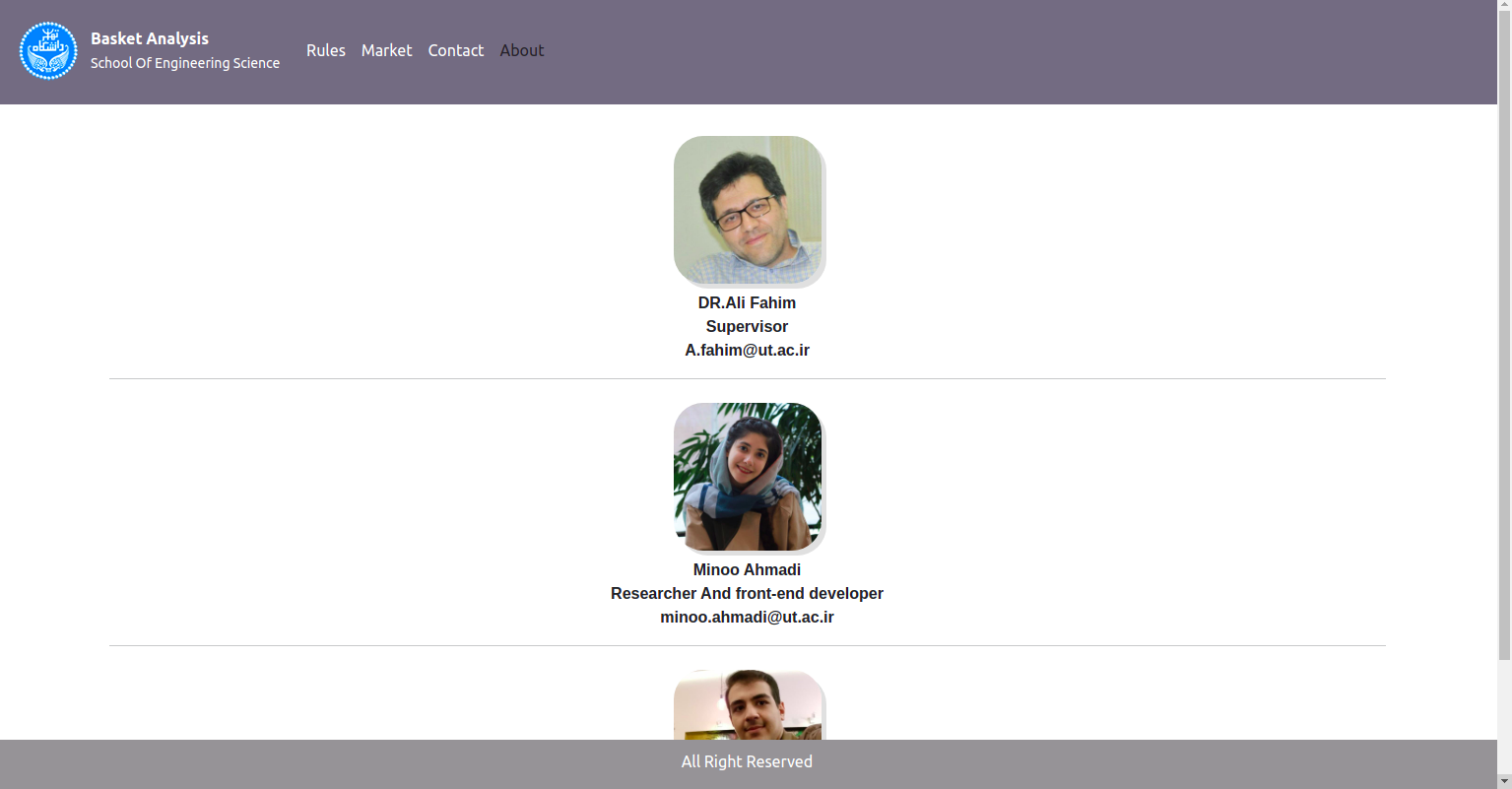
شکل 15– صفحه گرفتن قوائد

تصویر زیر نشان‌دهنده بخش انتخاب کالا و پیشنهاد گرفتن کالا یا مشتری است. ابتدا کالاهایی که می‌خواهیم در سبدمان باشد یا به مشتری‌هایی بفروشیم را انتخاب می‌کنیم.سپس اگر بخواهیم پیشنهاد کالا بگیریم، باید بر روی گزینه Get Suggestions فشار دهیم و اگر بخواهیم مشتری از سیستم پیشنهاد بگیریم، گزینه Get Customers را برمی‌گزینیم.

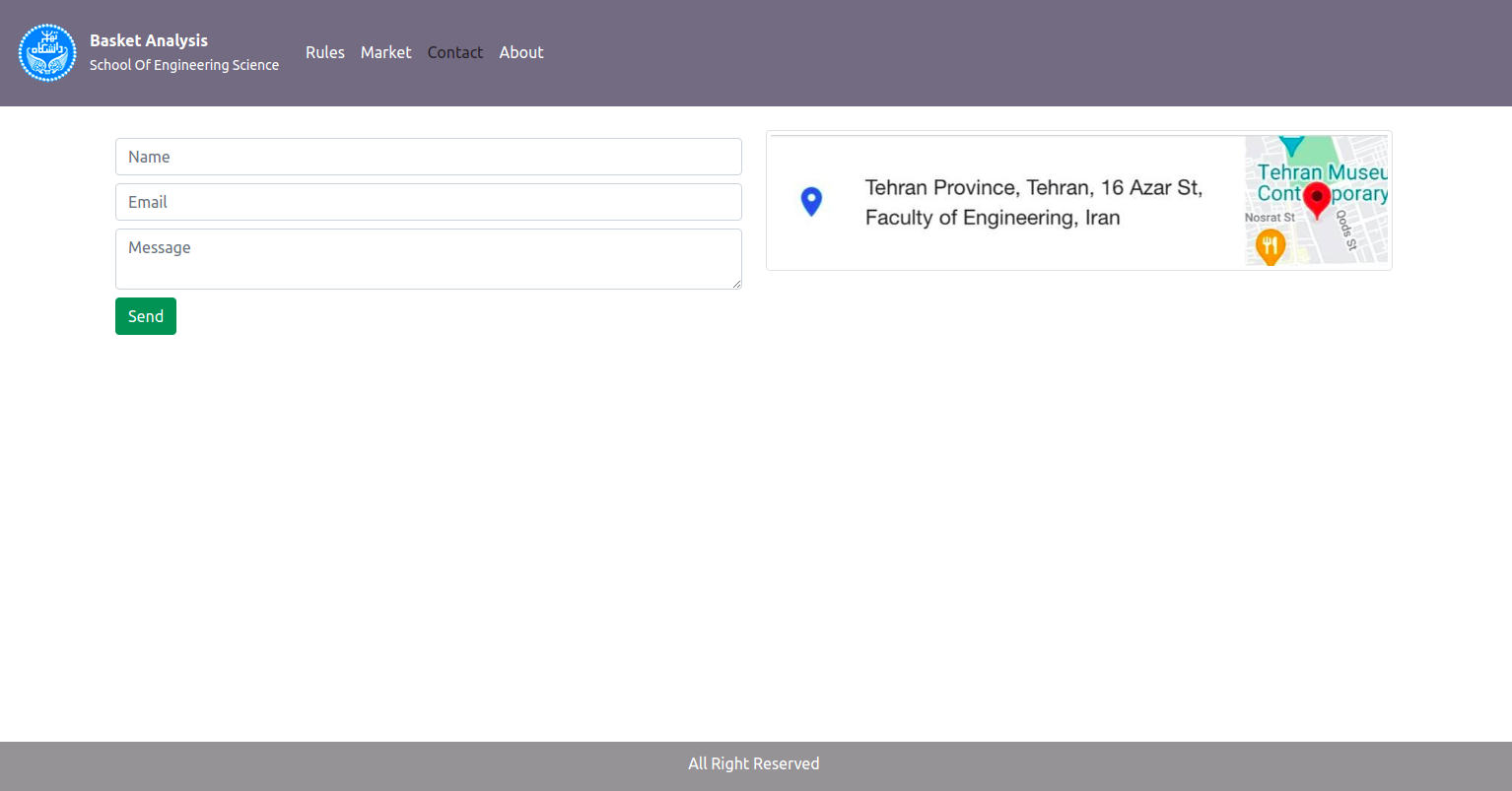


شکل 16- صفحه پیشنهادات

دو صفحه دیگر نیز برای معرفی افراد تیم و راه‌های ارتباطی با تیم ساخته شده است که در ادامه می‌بینید.



شکل 17– صفحه درباره ما



شکل 18- صفحه تماس با ما

* 1. اعتبار سنجی

از آن‌جایی که تعداد داده‌ها بالا است پس داده‌ها را به کشورهایشان محدود کردیم. ما برای آزمایش صحت این پروژه از دو روش استفاده کردیم. اولین روش چک کردن خروجی پیشنهادها با فاکتورهای فروشگاه بود. دومین روش چک کردن قوانین انجمنی به صورت دستی با معیارهای گفته شده در فصل 2 می‌باشد که نتیجه صحت کار را نشان می‌دهد.

1. بحث و نتیجه گیری
   1. مقدمه

با تحلیل و پیاده‌سازی الگوریتم‌های متفاوت به این نتیجه رسیدیم که الگوریتم Apriori در مقایسه با الگوریتم FP-Growth سرعت کمتر و حافظه اشغالی بیشتری دارد علت این موضوع به دلیل تفاوت در تولید یا عدم تولید کاندیداها و هم چنین تفاوت محاسبه فراوانی الگوهای تکرار شونده و تعداد دفعات پیمایش داده‌ها است.

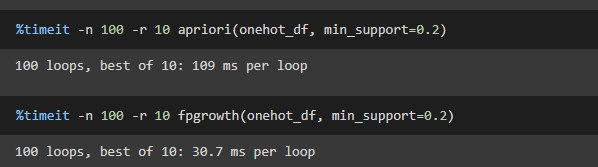
این نرم‌افزار نمونه مشابهی در ایران ندارد و می‌تواند به صاحبان کسب‌و‌کار در بازاریابی و فروش محصولات کمک نماید.

این الگوریتم‌ها بر روی داده‌های گرفته شده از دانشگاه UCI اجرا شده و نتایج قابل قبولی داده است. هم‌اکنون رضایت یکی از شرکت‌های ایرانی را نیز برای اخذ داده‌های آن‌ها و اجرای نرم‌افزار بر روی آن‌ها جلب کرده‌ایم که بزودی اجرا خواهد شد.

* 1. محتوا
     1. جمع‌بندي

با توجه به آنکه در پیاده‌سازی وب اپلیکیشن دو راه برای خواندن داده‌ها انجام گرفته شده است، با آزمایش هر دو روش بر روی داده‌های مختلف این نتیجه حاصل شد که روش خوانش داده از پایگاه داده سریع‌تر از خوانش داده از فایل csv است.

همچنین در پروژه‌ی خود برای بالا بردن سرعت اجرا از الگوریتم FP-Growth به جای الگوریتم اپریوری استفاده کردیم که در تصویر زیر نتیجه‌ی مقایسه‌ی دو الگوریتم با داده‌های یکسان در ۱۰۰ بار اجرا مشاهده می‌شود.



شکل 19- مقایسه الگوریتم‌های پروژه

با کمک از این نرم‌افزار شرکت‌های هدف می‌توانند فروش خود را بهبود ببخشند. همچنین می‌توانند در انبار خود کالاها را به صورت بهینه ذخیره کنند و کالاهایی که مشتریان ممکن است در آیندهبه آن‌ها نیاز داشته باشند را تهیه کنند.

* + 1. نوآوري

بر خلاف پروژه‌ها و پژوهش‌های انجام گرفته تا به امروز که اکثرا تنها با استفاده از یک الگوریتم frequent pattern mining را انجام می‌دادند ما این کار را با به کارگیری چندین الگوریتم متفاوت انجام دادیم تا نتایج حاصل از آن‌ها را باهم مقایسه کنیم.

در محاسبه قوانین انجمنی در پروژه جهت بالا بردن دقت نتایج نهایی کار، قوانین را بر اساس معیارهای گفته شده در فصل ۲ هرس نمودیم.

در آخر لازم به ذکر است که با توجه به سخت بودن استفاده از الگوریتم‌ها برای شرکت‌ها ما تلاش کردیم تا برای اولین بار در ایران بستری را فراهم سازیم تا شرکت‌ها با کمک وب اپلیکیشن به راحتی بتوانند با چند کلیک ساده نتایج را استخراج کنند، در واقع با پیاده سازی الگوریتم‌های پیچیده در بستری که استفاده آن برای کاربران بسیار آسان است، دسترسی شرکت‌ها را به این نتایج و پیشنهادات ممکن ساختیم.

* + 1. پيشنهادها

پیشنهاد می‌شود این پروژه در شرکت‌های داخلی پیاده‌سازی شود و نتایج آن (که شامل درصد بهبودی هزینه‌های جاری شرکت، افزایش فروش و ...) به صورت کمی منتشر شود. همینطور اینکار باعث می‌شود که پارامترهای ما هم دقیق‌تر تنظیم شوند.

* + 1. محدودیتها

بدلیل کاهش هزینه از داده‌های خارجی موجود در اینترنت استفاده کردیم. همچنین این نرم‌افزار نیاز به راه اندازی بر روی یک سرور قوی است ولی ما بر روی سیستم‌های خانگی آن را تست کردیم.

1. مراجع

مراجع

[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. 032219. 10.1088/1755-1315/252/3/032219.

[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&#x27;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] Ghorbani Kh., Khalili Ali, Alavipanah S.K., Nakhaeizadeh Gh.R., “ Comparative study of the meteorological drought indices (SPI and SIAP) using data mining method ( Case study of Kermanshah Province)”, Journal of water and soil (Agricultural science and technology), Aug-Sep 2010, vol. 24, number3, p. 417-426.

[7] Raorane, A.A, Kulkarni R.V., and Jitkar B.D. (2012). Association Rule – extracting Knowledge Using Market Basket Analysis. Research Journal of Recent Sciences, 01(2): 19-27.

[8] Annie M.C., Loraine Ch. and Ashok Kumar D. (2012). Market Basket Analysis for a Supermarket based on Frequent Itemset Mining. International Journal of Computer Science, 9(3): 257-264.

[9] Bottcher, M. Spott, M. Nauck, D., and Kruse, R. (2009). Mining changing customer segments in dynamic markets. Expert Systems with Applications, 36: 155– 164.

[10] Gupta, A. K., and Gupta, Ch. (2010). Analyzing Customer Behavior Using Data Mininghniquse: Optimizing relationships with customer. Management Insight, VI (1): 92-98.

[11] Miguéis, V.L., Poel, Dirk Van den, Camanho, A.S., and Cunha, Joo Falco e (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. Expert Systems with Applications, 39: 11250–11256.

[12] Ngai, E.W.T., Xiu, Li and Chau, D.C.K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36: 2592–2602.

[13] Ramageri, Bharati M., and Desai, B.L. (2013). Role of data mining in retail sector. International Journal on Computer Science and Engineering (IJCSE), 5(01): 47-50.

[14] غفاری آشتیانی، پیمان و داودی، معصومه، 1395، بررسی و تحلیل سبد بازار و چیدمان محصولات فروشگاه‌های زنجیره‌ای، <https://civilica.com/doc/916325>

[15] تارخ، محمدجعفر و شريفيان، كبري (۱۳۸۹). كاربرد داده‌كاوي در بهبود مديريت ارتباط با مشتري. فصلنامه علمي پژوهشي مطالعات مديريت صنعتي، ۶(۱۷): ۱۵۳-۱۸۱.

[16] فشاركي، مهدي؛ شيرازي، حسين و بخشي، اكرم (۱۳۸۹). استخراج دانش از پايگاه داده نرم‌افزارهاي مديريت اطلاعات و مستندات با تكنيك‌هاي داده‌كاوي. فصلنامه علمي پژوهشي پژوهشگاه علوم و فناوري اطلاعات ايران، ۲۶(۲): ۲۵۹-۲۸۳.

[17] ديوسالار، علي و بهبود، وحيد (1385). مديريت بر اطلاعات مشتري در مديريت زنجيره تأمين با استفاده از رويكرد داده كاوي. فصلنامه اطلاع رساني، آموزشي و پژوهشي مديريت فردا، ۴(۱۴ و ۱۳): ۷۳-۶۷.

[18] تقوي‌فرد، محمدتقي و خواجوند، سمانه (۱۳۹۲). بخش‌بندي خوشه‌اي مشتريان بانك (مورد

مطالعه: بانك صادرات ايران). دو فصلنامه علمي-پژوهشي كاوش‌هاي مديريت بازرگانی، ۵(۹): ۳۹-۶۴.

**Abstract:**

In this project, the goal is to launch a software to optimize and improve the sales process of companies. In this system, using data mining algorithms and creating rules based on company invoices, we tried to give practical and accurate suggestions to business owners so that they can use these suggestions to make their marketing more accurate and efficient. These suggestions consist of two parts. The first type of offer is based on the customer's current shopping cart, which gives the goods related to the cart to the employer so that the employer can increase the probability of selling them by offering those goods. The second type of offer in the system is based on previous purchases of customers, which offers the employer, based on the goods of his choice, potential customers who want those goods. It should be noted that due to the difficulty of using algorithms for companies, we tried to provide a platform for the first time in Iran so that companies can easily extract results with the help of a software with a few simple clicks, in fact by implementing Sophisticated algorithms in a platform that is very easy for users to use, we made it possible for companies to access these results and suggestions.

Flask restful tool has been used to build this software for back-end development. The Vue-JS tool is also used for front-end development. The whole system data is stored in a database with Postgresql management system. At the core of this system, Apriori and FP-Growth algorithms are used, which have their positive and negative points in the report.

**Keywords:**

Data Mining, Association Rules, Frequent Pattern mining, Basket Analysis

|  |  |  |  |
| --- | --- | --- | --- |
| دانشگاه تهران - ویکی‌پدیا، دانشنامهٔ آزاد | University of Tehran  College of Engineering    Faculty of Engineering Science | Description: Related image | |
| **Frequent Pattern Mining** | | | |
| A thesis submitted to department  In partial fulfillment of the requirements for  The degree of Bachelor in  Engineering Science | | | |
| **By:**  **Minoo Ahmadi**  **Supervisor:**  **Dr. Ali Fahim**  **Advisor:**  **Dr. Ali Fahim** | | | |

1. https://archive.ics.uci.edu/ml/datasets/online+retail [↑](#footnote-ref-1)