**توصیه مکان در شبکه های اجتماعي مبتني بر موقعیت بر اساس تحلیل سری های زماني حضور كاربران**

در سال­های اخیر با توسعه و پیشرفت سریع وسائل سیار و دسترسی به اینترنت در همه­جا، سرویس­های شبکه­های اجتماعی مانند Facebook، Twitter، Yelp، Foursquare و غیره، به معروفیت قابل­توجهی دست یافته­اندبا پیشرفت سریع تکنیک­های مکان­یابی مانند GPS و Wi-Fi که در تجهیزات همراه هوشمند به طور عادی استفاده می­شوند و کاربران می­توانند مکان­های فیزیکی خود را در شبکه­های اجتماعی گزارش دهند. شبکه­های اجتماعی با رکوردهای اضافه از مکان جغرافیایی کاربران، شبکه­های اجتماعی مبتنی بر مکان نامیده می­شوند. بواسطه­ی کمک گرفت از وسائل همراه و شبکه­های اجتماعی مبتنی بر مکان آنلاین، کاربران قادر به اشتراک گذاشتن اطلاعات مربوط به موقعیت جغرافیایی با مراجعه­های مربوطه خواهند بود. چنین اطلاعاتی سبب پیش آمدن فرصت­ها و چالش­ها برای سیستم­های توصیه­گر می­شوند.

سيستم توصيه­گري که براساس تاريخچه مكان­هاي بازديد شده توسط کاربر و نيز ساير اطلاعات موجود در شبكه اجتماعي، مكان هاي بازديد نشده را به کاربر پيشنهاد مي دهد، سيستم توصيه­گر در شبكه­هاي اجتماعي مبتني بر مكان ناميده مي­شود.سيستم­هاي توصيه­گر همواره با چالش­ها و مشكلاتي مواجه هستند. شروع سرد و پراکندگي داده­ها از اساسي­ترين مشكلات اين سيستم­ها مي­باشند. هنگامي که کاربري براي اولين بار وارد يك شبكه اجتماعي مي­شود، چون از کاربر هيچ گونه اطلاعاتي وجود ندارد، سيستم توصيه­گر با مشكل شروع سرد روبه رو مي­شود. پراکندگي داده­ها بدين معني است که اطلاعات در سيستم وجود دارد اما پراکنده هستند و نمي­توان به درستي و با قطعيت گفت که چه آيتمي مقبوليت بيشتري دارد. علاوه بر اين مشكلات، براي يك توصيه مكان خوب بايد به غير از مكان جغرافيايي، زمان مراجعه، دوستي کاربران و دسته بندي­هاي نقاط مورد علاقه را در توصيه مكان در نظر گرفت.از ميان عوامل موثر در توصيه مكان، زمان مراجعه اهميت بيشتري دارد زيرا رفتار ورود به سيستم کاربر ممكن است به صورت دوره­اي و وابسته به زمان باشد، به عنوان مثال چك کردن)ورود به سيستم(در نزديكي کار در صبح و ورود در عصر نزديك به خانه.

در پروژه پيش رو يك سيستم توصيه­گر ارائه مي­شود که يافتن مكان­هاي مورد علاقه را در بستر زمان توصيه مي­کند. اين سيستم از سري­هاي زماني داده­هاي مكان براي پيش­بيني استفاده مي­کند. در اين روش مراجعات کاربران شبكه هاي اجتماعي به مكان ها در بازه هاي زماني يكسان مانند ساعت، روز و فصل ثبت مي­شود و از اين داده­ها براي توصيه­هاي آينده در بازه­هاي زماني مشابه استفاده مي­گردد. در اين پژوهش از سري­هاي زماني به عنوان يك الگو براي پيش­بيني مكان در آينده استفاده مي­شود مانند پيش­بيني مصرف انرژي برق براي فصل­هاي سال و ساعت­هاي يك روز مكان­ها براي سال­ها و روزهاي آينده بر اساس داده­هاي موجود. اين مدل مي­تواند مشكل شروع سرد و پراکندگي داده­ها را به طور قابل­توجهي کاهش دهد و با استفاده از اطلاعات زماني، منجر به افزايش دقت توصيه ها گردد .

LBSN به عنوان نرم­افزار محبوب خدمات مبتنی بر مکان، فرصت بی­سابقه­ای برای مطالعه­ی رفتارهای حرکتی و سیار انسانی برای توصیه­های POI در جنبه­های مکانی، زمانی، اجتماعی و محتوایی فرآهم می­سازد. در مقایسه با توسعه­ی سنتی سیستم­های توصیه­گر، توسعه­ی سیستم توصیه­گر POI بسیار پیچیده است. توصیه­های POI بنا به برخی دلایل به چالش کشیده می­شود:

1. ماتریس­های مراجعه­ی کاربران POI بسیار پراکنده است چرا که کاربران تنها بخش کوچکی از POIها را در LBSN مشاهده می­کنند و بنابراین روش­های توصیه دارای مشکل کمبود اطلاعات است.
2. علایق کاربران در توصیه­های POI می­تواند به طرز چشمگیری در زمان­ها و مکان­های مختلف متنوع باشد.
3. انواع مختلف اطلاعات بافت در توصیه­های POI را نظیر مختصات جغرافیایی POI، نوع زمان مراجعه، دوستی کاربران، دسته­بندی­های POI و غیره می­توان بدست آورد. با این حال، برخلاف توصیه­های سنتی، اطلاعات متنی مرتبط با POI معمولا ناقص و مبهم است.

تجزيه و تحليل و پيش­بيني سري زماني در داده­هاي سري زماني، در بسياري از کاربردهاي مانند کسب و کار، بازار سهام و تبادل، آب و هوا، تقاضاي برق، هزينه و استفاده از محصولات مانند سوخت، برق، و در هر جايي که داراي تغييرات خاص فصلي يا مرسوم در طول زمان باشد، اهميت پيدا مي­کند. پيش بيني داده هاي سري زماني، سازمان را با اطلاعات مفيدي که براي تصميم گيري هاي مهم لازم است، مواجه مي کند. با استخراج اطلاعات و الگوهاي مفيد با استفاده از داده کاوي، اين الگوها را مي­توان براي پيش­بيني آينده تحليل کرد.

سيستم توصيه­گر مكان پيشنهادي با بهره­گيري از داده­هاي سري­هاي زماني در شبكه­هاي اجتماعي مبتني بر مكان با توجه به معيار زمان يك روش مناسب براي حل مشكل شروع سرد و پراکندگي داده­ها فراهم مي­کند و دقت توصيه­ها در اين سيستم افزايش مي­يابد.

* پيشنهاد مكان به کاربران شبكه­هاي اجتماعي مبتني بر مكان با بالاترين دقت
* پيشنهاد مكان با تكيه بر سابقه بازديد مكان ها توسط کاربران
* ارائه يك سيستم توصيه­گر مكان در شبكه­هاي اجتماعي مبتني بر مكان با دقت بالا با بهره­گيري از سري­هاي زماني و داده کاوي بر اساس زمان حضور افراد در مكان­ها
* استفاده از داده­هاي سري­هاي زماني جهت حل مشكل شروع سرد و پراکندگي داده­ها

**سری های زمانی**

یک سری زمانی مجموعه از مشاهداتی است که برحسب زمان مرتب شده­اند. به عبارت دیگر می­توان گفت یک سری زمانی عبارت است از داده­هایی که از مشاهده پدیده در طول زمان بدست آمده­اند. یک سری زمانی دنباله­ای از مشاهدات بر روی یک متغییر مورد توجه در نظر می­گیریم. متغییر در نقاط گسسته­ای از زمان که معمولا فاصله­های مساوی، دارند مشاهده می­شود. تجزیه و تحلیل سری­های زمانی متضمن توصیف فرآیند یا پدیده­ای است که تولید دنباله می­­کند. در یک تقسیم بندی کلی سری­های زمانی را به پیوسته و گسسته تقسیم می­کنیم. یک سری زمانی را پیوسته گوییم هرگاه مشاهدات بطور پیوسته از زمانی ایجاده شده باشد. و یک سری زمانی را گسسته گوییم هرگاه مشاهدات فقط در زمان­های معینی که معمولا به فواصل مساوی از یکدیگرقرار دارند اخذ شده باشد. و هم چنین یک سری زمانی را تصادفی می­گوییم اگر پیش بینی کامل آن غیر ممکن باشد. و آنرا غیر تصادفی گوییم اگر بتوان آن را کامل پیش بینی کرد(پاولوس،2017).

طبقه بندی انواع پیش بینی با استفاده از داده های سری زمانی به شرح زیر است.

طبقه بندی زمانی: این فرآیند که کلاس داده ها را از هم جدا می­کند(با اعمال طبقه بندی).

پیش بینی زمانی: این فرآیند ارزش های شناخته نشده را از تاریخچه از داده های را پیدا می­کند.(مانند عمل داده کاوی که از داده های که در ابتدا ارزش شناخته شده ای ندارد با آنها برای آیند سازمان پس از اعمال داده کاوی بر روی داده­های بی ارزش می توان به نتایج بسیار ارزشمندی برای آینده سازمان را یافت کرد).

خوشه بندی زمانی: این فرآیند گروه بندی داده­های مشخص را از پایگاه داده زمانی را مشخص می­کند.

رگرسیون زمانی: این فرآیند تلاش بر این دارد که عملیاتی را روی داده­ها انجام دهد که آنها را با کم ترین خطا مدل سازی کند.

خلاصه سازی زمانی: این فرآیند داده های زمانی را خلاصه می­کند.

کشف داده­های خارج ازخط زمانی: این فرآيند تشخیص داده های عمومی که با دیگر داده های پایگاه داده هم­خوانی ندارد.

پیش­بینی سری­های زمانی در بسیاری از کاربردهای که با داده­های عددی سرو کار دارند استفاده می­شود. پیش­بینی بر اساس سه فضای زمانی کاربرد دارد.

دوره زمانی کوتاه مدت، دوره زمانی متوسط، دوره زمانی بلند مدت

در پیش بینی کوتاه مدت بر روی چارچوب کم تر از سه ماه را مدنظر دارد. و پیش بینی زمان متوسط معولا بین سه ماه تا یک سال گفته می­شود و دوره زمانی بلند مدت بیش از یکسال را تمرکز دارد.

توصیف رفتار یک سری زمانی، اجزاء زیر را برای سری زمانی در نظر می توان گرفت.

روند یا تمایل بلند مدت[[1]](#footnote-1)

تغییرات دوره­­ای[[2]](#footnote-2)

تغییرات فصلی[[3]](#footnote-3)

تغییرات نامنظم[[4]](#footnote-4)

**روند**

روند یا تمایل بلند مدت عبارت است از تحول متغییر مورد مطالعه در یک دوره طولانی بدون در نظر گرفتن تغییرات دوره­ای، فصلی و نامنظم به عبارت دیگر می­توان گفت روند عبارت است از حرکات رو به بالا و پایین یک سری زمانی که نشان دهنده کاهش یا افزایش بلند مدت یک سری زمانی است.

**تغییرات دوره­ای یا سیکل**

تغییرات دوره­­ای عبارت است از تکرار حرکات روبه بالا و پایین حول سطوح روند.این نوع تغییرات دارای نوسان بیشتر از یک سال می­­باشند. نوسانات دوره­ای ممکن است دقیقا از طرح های مشابهی بعد از فواصل زمانی مساوی پیروی کنند. ولی همیشه این طور نیست یک دوره کامل را که معمولا هفت تا نه سال طول می­­­­کشد.اصطلاحات یک دوره می­نامند. یکی از معمولی ترین نوسنات سیکلی داده­*­*­های سری زمانی، سیکل تجاری است. سیکل تجاری وقوع مکرر دوره­های رونق و رکود است.

**تغییرات فصلی**

تغییرات فصلی تغییراتی هستند که دوره­های تناوبی کوتاه پیش می­آیند. این تغییرات مربوط به عواملی هستند که به طریقی منظم و چرخه­ای روی یک دوره کمتر از یک سال عمر می­کنند. در واقع تغییرات فصلی رفتار دوره­ای متغییر را نشان می­دهد. یعنی رفتاری که معمولا هر سال در همان فصل تقریبا با همان شدت روی می­دهد.

**تغییرات نامنظم**

تغییرات نامنظم عبارت است از حرکات پراکنده در یک سری زمانی که از الگوی منظم و مشخصی پیروی نمی­­کنند. در واقعاین حرکات بیان می­کنند. که پس از محاسبه روند، تغییرات دوره وتغییرات فصلی چه چیز دیگری در سری زمانی بجا می­ماند.

نوسانات نامنظم معمولا ناشی از وقایع غیر معمولی هستند. که قابل پیش نیستند.

**مدل سازی سری­های زمانی**

امروزه سری های زمانی در بازار سهام و انتخاب سهام در سرمایه گزاری و تصمیم­گیری برای آینده حیاتی هستند. پیش بینی بازار آینده بازار سهام با استفاده از داه های گذشته بازار سهام با استفاده از رویکرد محاسبات نرم برای پیش بینی داده­های مالی بازار سهام و سری­های زمانی برای جلوگیری از کاهش ریسک مالی و سرمایه گذاری برای آينده مالی بازار سزی های زمانی بسیار مورد استفاده­ قرار می گیرند.

در واقعه سری ­زمانی سعی برآن است که با بررسی گذشته سری، الگوی احتمالی مولد داده ها را شناسایی کرده و بر مبنای این الگوریتم درباره آینده سری اظهار نظر نماییم.

مدل احتمالی که به سری زمانی برازش داده می­شود. باید بتواند به نحوه مناسبی مشاهدات سری را مدل سازی کند. برای بررسی اینکه آیا یک مدل احتمالی واقعا بیان کننده داده ها است یا خیر، می­توان خطاهای پیش بینی را مور تجزیه و تحلیل قرار داد. اگر مدل به نحوه رضایت بخشی نماینده فرآیند باشد، آنگاه انتظار می­رود مقدار متوسط خطاهای پیش بینی نزدیک به صفر باشد.

**مدل سازی سری های زمانی فازی**

نظریه فازی به ارائه روش­های قابل پیاده سازی برای حل مسائل مختلف در محیط های نادقیق می­پردازد. علی ­رغم كاربردهاي گسترده كامپيوتر هاي ديجيتال در محاسبات و كنترل سيستم­ها، استفاده از آن ها درشرايطي كه اطلاعات دقيقي در مورد پارامترهاي مؤثر در سيستم وجود نداشته باشد عملا غيرممكن بود. با معرفي نظريه فازي توسط زاده اين مشكل در حد قابل تصوري حل شد به ويژه اين كه با ارايه اين نظريه امكان مدلسازي رفتار پيچيده انساني در مسايل مديريتي و حل بسياري از مسايل مربوط به مدلسازي رفتار غير خطي سيستم­ها فراهم آمد**.**

مجموعه هاي فازي در سال ۱۹۶۵ توسط پروفسور علي اصغر لطفي زاده كه در سطح جهاني به زاده شهرت دارد به عنوان يك امكان وسيله رياضي براي نمايش ابهام در مقادير كلامي مانند زياد، كم، تاحدودي و ... معرفي شد،ايده او در اصل تعميم نظريه مجموعه هاي كلاسيك است. از اواسط ۱۹۷۰ به بعد تئوريهاي مختلفي در رابطه با نظريه فازي توسط دانشمندان ايراني، اروپايي، آمريکايي و ژاپني ارائه شده است. ولي دانشمندان ژاپني تحقيقات قابل توجهي را در رابطه با استفاده عملي از نظريه فازي انجام داده اند. عليرغم كاربردهاي فراوان نظريه فازي در مسايل پيچيده اي كه با اطلاعات مبهم سروكار دارند، بايد توجه داشت كه سيستم­هاي فازي كارايي كمتري نسبت به الگوريتم هاي دقيق دارند**.** ولي نياز به دقت در سيستم­هاي فيزيكي، نيازبه سرمايه گذاري فراوان در دستيابي به اطلاعات دقيق و صرف وقت زياد براي توسعه مدلهاي دقيق را در پي دارد **.**

به عبارت ديگر در سيستم­هاي پيچيده، قيمت متناسب با زمان است. دستيابي به دقت عمل نه تنها هزينه زياد دارد بلكه مستلزم پيچيدگي بيشتراست. در عين حال در صورت مواجه با مسئله اي كه اطلاعات در مورد آن گنگ يا مبهم است امكان دسترسي به مدل دقيق وجود نداشته و استفاده از سيستم­هاي فازي امكانپذير ترين راه حل است .

نظريه احتمال براي پيش بيني نتيجه يك رويداد در آينده بكار مي­رود. رويدادي كه در آينده قرار است اتفاق بيافتد و نتيجه آن در حال حاضر مشخص نيست. در واقع نظريه احتمال به رويدادهاي تصادفي مرتبط مي باشد.

در حاليكه فازي به ”بي دقتي“ و مفاهيم نادقيق كه در زبان طبيعي بكار مي روند مرتبط است و هميشه با يك رويداد همراه نيست. در واقع نظريه فازی عدم قطعيت غير احتمالي را پشتيباني مي­كند.

در یک مجموعه فازی دارای تابع عضویت می­باشد. نحوه ایجاد مجموعه­های فازی و تعریف تابع عضویت آن­ها بستگی به دامنه وزمینه کاربری سری زمانی دارد. تعریف یک مجموعه فازی برای مفهوم مورد نظر با تعریف یک تابع عضویت برای آن کامل می­شود.تعریف تابع عضویت مناسب بسیار مهم است.زیرا اگر تابع عضویت تعریف شده برای مجموعه فازی مناسب نباشد کلیه تحلیل و بررسی­های پس از آن دچار انحراف می­شود.

مدل سازی سری زمانی با روش فازی و تابع عضویت آن می­توان صورت گیرید. در پژوهش حاضر بر روی سری زمانی با توصیه مبتنی بر مکان را می­توان از تابع عضویت رد مدل سازی استفاده کرد.

انواع تابع عضویت

تابع عضویت استاندارد

تابع عضویت ذوزنقه­ای

تابع عضویت گوسی

تابع عضویت زنگوله­ای

تابع عضویت سیگمویدال

توابع عضویتی که برای مدل سازی سری های زمانی در مبحث فازی می­توان استفاده کرد.

**الگوریتم سری‌های زمانی**

روش سری‌های زمانی یکی دیگر از روش‌های پیش بینی است. این مدل نوعی الگوریتم رگرسیون است که برای پیش بینی مقادیر پیوسته مانند فروش محصولات در طول زمان استفاده می‌شود. سایر الگوریتم‌های ماکروسافت مانند درخت تصمیم نیازمند ستون‌های اضافی از اطلاعات به عنوان ورودی برای روند پیش بینی می باشند در حالی که در الگوریتم سری زمانی نیازی به این ستون‌های اضافی نمی باشد. در این الگوریتم می توان از خروجی یک سری، بر پایه رفتار سری دیگر استفاده نمود. در واقع سری‌های زمانی، مجموعه‌ای اطلاعات از افزایش‌های متوالی داده‌ها که در یک دوره زمانی جمع آوری شده اند می‌باشد. دنیای پیرامون ما ثابت نیست و متغیرهای بسیاری با تغییر زمان ارزش خود را تغییر می‌دهند، ترتیب ارزش‌های یک متغیر درطول زمان،یک سری زمانی را تشکیل می‌دهد. هدف اصلی از جمع آوری داده‌های سری زمانی پیش‌بینی و یا پیشگویی درباره مقادیر آینده است.

**مدل سازی سری­های زمانی شبکه عصبی**

شبکه های عصبی مصنوعی که معمولاً به عنوان" شبکه های عصبی" نام برده می شوند یک الگوی ریاضی مبنی بر سیستم زیستی است، سیستم های عصبی یک الگوریتم برای بهینه سازی و یادگیری آزادانه بر اساس مفاهیم الهام گرفته از تحقیق در ماهیت مغز می باشند. استفاده از قابلیتی شناخته شده به عنوان نورون ها اجزاء ساختاری خود را سازماندهی می کند، در نتیجه محاسبات معینی را بسیار سریع تر از کامپیوتر دیجیتال انجام می دهد. در حالت کلی شبکه عصبی ماشینی است که طراحی شده تا روشی مشابه با کاری که مغز برای انجام وظایف خاص یا عملکرد قابل توجه بر اساس مطالعات دکتر انجام می دهد را مدل سازی کند ."شبکه عصبی یک پردازنده توزیع شده موازی بزرگی است که از واحد های پردازش ساده ساخته شده است، و دارای یک تمایل طبیعی برای ذخیره سازی دانش تجربی و ایجاد دسترسی به آن لایه ورودی:این لایه ورودی هارا دریافت می کند و برحسب قدرت ارتباطش با لایه بعد سیگنال ورودی را به لایه بعد می فرستد. لایه میانی:تعداد لایه های میانی و تعداد نورون های آن دلخواه است.لایه های میانی بایدبا دقت انتخاب شوند تا خروجی مناسبرا به مابدهند. لایه خروجی: گروه دیگری از نورو نها نیز از طریق خروجی های خود، جهان خارج را می سازند.

شبکه عصبی مصنوعی مانند یک تابع عمل می کند که به تعداد نورون های ورودی،ورودی می گیرد و به تعداد نورون های خروجی،خروجی می دهد.

مدل پروژه حاضر از روش مدل خطی آریما در محیط نرم افزاری متلب استفاده خواهیم کرد.

مدل آریما، مدلی است که سری­های زمانی (وقایع و رخداد­های هستند که در یک زمان متوالی مشخص مثل هفتگی و ماهانه یا سالانه برداشت کرده­ایم) مدل آریما کاربردهای فراوانی دارد مانند:تخمین پارامترها،برای پیش­بینی پارامترها،مدل سازی سری­های زمانی وغیره موارد هم استفاده می­شوند، انعطاف زیادی که مدل آریما دارد به دلیل انعطاف پذیری پارامترهای ریاضی مدل آریما است. مدل آریما برای پیش­بینی­سری­های زمانی توصیه مکان در شبکه­های اجتماعی باتوجه به حضور کاربران را در محیط نرم افزاری متلب پیاده سازی خواهیم کرد.

جمع­آوری داده،اصلاح داده ،بررسی­داده اعمالی بر روی مدل آریما

آموزش مدل آریما

جمع­آوری داده­ها

پیش­بینی مدل آریما

ایجاد مدل آریما بدون هیچ گونه پارامتر و یا ضرایبی نامشخصی

پارامتر دهی به مدل آریما (پارامترهای مطلوب) به مدل آریما پژوهش را خواهیم داد

در این مرحله آموزش مدل آریما با داده­های ما انطباق پیدا می­کند.

با رسم نموداردر نرم افزار متلب پیش­بینی رو انجام خواهیم داد

توصیف آماری مدل آریما

در پاراگراف قبل از متغیر Y**t** به عنوان متغیر مدل آریما از آن صحبت شددر پاراگراف های بعد از آن متفیر به بررسی فرمول های آماری از مدل آریما از جزترین مقداری که مدل آریما تحقیق حاضر خواهد گرفت تا انواع مدل سازی به روش آریما خواهیم پرداخت

* متغير تصادفي ايستاي ضعيف **Y**t را در نظر بگيريد
* اميد رياضي متغير تصادفي را**Y**t به شرح زیراست:

* واريانس متغير تصادفي **Yt** به شرح زیر است:

* تابع اتو كوواريانس متغير تصادفيY**t و** Y**t-k** به شرح زیر است **:**
* تابع خود همبستگي (ACF) بين دو متغير تصادفيY**t و** Yt-k **:**
* تابع خود هم­بستگي جزئي (PACF) بين دو متغير تصادفي Yt و Yt-k **:**

**تعريف: عبارت است از هم­بستگي بين دو متغير تصادفي** Yt و Yt-k **پس از حذف وابستگي خطي منحصر بفرد آنها با مقادير بينابيني** Yt-1 ، Yt-2 ، Yt-3 ، ...، Yt-k+1

Zt-k

Zt-k+1

Zt-3

Zt-2

Zt-1

Zt

به بستر فرمول ریاضی به شرح زیر است:

**مدل‌هاي اتو رگرسيو[[5]](#footnote-5)**

ساده‌ترين نوع مدل‌هاي پيوسته‌اند كه بر مبناي مدل‌هاي زنجيره ماركوف و توماس فیرینگ[[6]](#footnote-6) توسعه يافته‌اند. به طوركلي به دو دسته زیر قابل تقسيم‌اند.مدل‌هاي AR با پارامترهاي ثابت كه براي مدل‌سازي سري‌هاي زماني سالانه رايج مي‌باشند، مدل‌هاي AR با پارامترهاي متغير در زمان كه براي مدل‌سازي سري‌هاي زماني پريوديك (براي مثال ماهانه و فصلي) به كار مي‌روند.اين مدل‌ها هم­بستگي بين توالي زماني متغيرها را در نظر مي‌گيرند. به عبارت ديگر، مقدار متغير در يك زمان معين بستگي به مقدار آن متغير در زمان‌هاي قبلي دارد، فرض اوليه اين مدل‌ها نرمال بودن سري زماني داده‌ها است.

مدل‌هاي اتو رگرسيو با پارامترهاي ثابت

فرم پايه مدل‌هاي اتورگرسيو مرتبه p با پارامترهاي ثابت براي متغير نرمال استاندارد شده Yt به صورت زير است.

* تابع خودهم­بستگي­مدل‌هاياتو رگرسيو با پارامترهاي ثابت به شرح زیر می باشد.



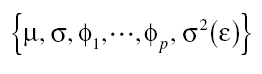
* تابع خودهم­بستگي جزئي








مجموعه پارامترهاي يك مدل اتو رگرسيو با پارامترهاي ثابت یا به اختصار AR(p)

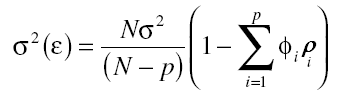


که برآورد پارامترهاي مدل AR(p) به شرح زیر است:

ميانگين (μ) و واريانس (σ2)سري زماني متغير تبديل نشده (Xt)

ضرايب مدل (φi): براي اين منظور، ابتدا ضرايب خودهمبستگي سري زماني را در گام‌هاي تاخير مختلف محاسبه نموده سپس از معادلات ارائه شده در مورد PACF مدل AR براي تعيين پارامترهاي مدل AR استفاده مي‌كنيم. در مثال‌هايي از مدل AR كه در بخش‌هاي بعد بيان مي‌شود، نحوه محاسبه پارامترها تشريح خواهد شد.

واريانس خطا (σ2(ε))



**شرط ايستايي پارامترهاي مدل اتو رگرسيو با پارامترهاي ثابت**

پس از محاسبه پارامترهاي مدل AR(p) لازم است ايستايي پارامترها مورد بررسي قرار گيرد. براي اين منظور، كافي است كه ريشه‌هاي معادله مشخصه زير در دايره‌اي به شعاع واحد قرار گيرد:

چناچه مقدارui از یک کوچک تر باشد ایستا می­باشد.

* تعيين رابطه تابع خودهم بستگی





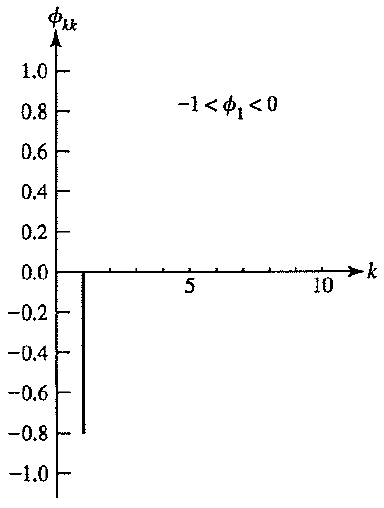
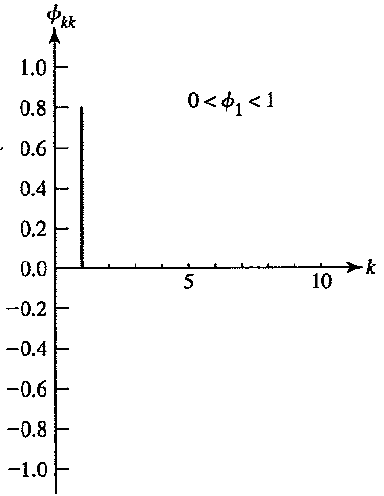




* تابع خود هم­بستگي جزئي مدل اتو رگرسيو





نمودار تابع خودهم­بستگی

**مدل‌هاي اتو رگرسيو با پارامترهاي متغير يا پريوديك**

فرم پايه مدل‌هاي اتورگرسيو مرتبه p با پارامترهاي پريوديك براي متغير نرمال استاندارد شده Yυ,τ به صورت زير است:



* **نماي جدول داده‌هاي مورد استفاده براي مدل AR(p) پريوديك**

جدول فصلی اتو رگرسيو

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Season  Year | 1 | 2 | 3 | (Box, G.E.P. and G.M. Jenkins 2016) … |
| 1990 | Z1990,1 | Z1990,2 | Z1990,3 | . |
| 1991 | Z1991,1 | . | . | . |
| 1992 | Z1992,1 | . | . | . |
| …. | . | . | . | . |
| µτ | µ1 | µ2 | µ3 | . |
| στ | σ1 | σ2 | σ3 | . |

**مدل‌هاي ميانگين متحرك [[7]](#footnote-7)**

تابع اتوكوواريانس مدلMA(q) **:** با فرض E(Zt)=0 ، و ضرب Zt-k در طرفين معادله MA(q) و گرفتن اميد رياضي از طرفين داريم (k گام تاخير است):

**واريانس مدل** MA(q)برابر است با:



**تابع خودهم­بستگي** **مدل** MA(q): از تقسيم γk بر γ0 بدست مي‌آيد:



**شرط پارامترهاي مدل** MA(q)**- شرط وارون‌پذيري** (Invertibility): شرط وارون‌پذيري بيان مي‌كند كه هر مدل MA(q) را مي‌توان به صورت يك مدل AR(p) نوشت اگر <1|φi| باشد يا ريشه‌هاي معادله مشخصه زير در دايره‌اي به شعاع واحد قرار گيرد:

**مدل‌هاي اتورگرسيو - ميانگين متحرك[[8]](#footnote-8)**

با توجه به معادله ACF مدل ARMA(p,q) ملاحظه مي‌شود كه ACF به ازاي گام‌هاي تاخير 1 تا q هم به ضرايب AR و هم به ضرايب MA وابسته است اما پس از گام تاخير q صرفاً به ضرايب مدل AR بستگي مي‌يابد،نمودار ACF مدل ARMA(p,q) به این صورت است. در صورتي كه q-p≥0 باشد شكل ACF تا q-p+1 گام تاخير حالت نامنظم داشته و پس از آن به صورت نمايي كاهش يافته يا به صورت موجي ميرا مي‌شود. اگر q-p<0 باشد، ACF به صورت نمايي كاهش يافته يا به صورت موجي ميرا مي‌شود.

تابع خودهم­بستگي جزئي (PACF) مدل ARMA(p,q): قبلاً مشخص شد كه PACF يك مدل AR(p) در تاخير p متوقف مي‌شود. در مقابل، PACF يك مدل MA(q) نامتناهي است و در تاخير q متوقف نمي‌شود بلكه پس از آن نيز ادامه يافته و به صورت موجي ميرا مي‌شود يا به صورت نمايي زوال مي‌يابد

شرايط پارامترهاي مدل ARMA(p,q):

شرط ايستايي: ريشه‌هاي معادله مشخصه مدل AR(p) در دايره‌اي به شعاع واحد قرار گيرد.

شرط وارون‌پذيري: ريشه‌هاي معادله مشخصه MA(q) در دايره‌اي به شعاع واحد قرار گيرد

شرايط استفاده از مدل‌هاي آرما[[9]](#footnote-9): عدم انحراف ظاهري از حالت ايستايي در سري زماني، تنزل سريع در تابع ACF سري زماني؛ چنانچه دو شرط فوق توسط سري زماني تامين نشود، مي بايست با يک تبديل مناسب آن دو را ارضا نمود. براي اين منظور مي توان از تبديل تفاضل (Difference) استفاده کرد و مدل حاصله آریما [[10]](#footnote-10)ناميده مي شود که داراي دو فرم است

فرم غيرفصلي مدل ARIMA از مرتبه (p,d,q) براي متغير نرمال استاندارد شده Yt به صورت زير است

که در آن φ(B) و θ(B) به ترتيب چند جمله اي از درجه p و q مي باشند:



و d مرتبه تفاضل را نشان مي دهد.

1. فرم فصلي مدل‌ ARIMA از مرتبه غيرفصلي (p,d,q) و مرتبه فصلي w(P,D,Q) براي متغير نرمال استاندارد شده Yt به صورت زير است:



فرم بسط يافته فصلي مدل‌ ARIMA(p,d,q)(P,D,Q)w

اعمال ضرایب و پارامترهای مدل آریما

طبق جدول زیر ما پارامترها و ضرایب مدل آریما را در نرم افزار متلب می دهیم و پس از آن عمل پیش­بینی را انجام خواهیم داد.



### داده های مدل سازی پروژه

### اولین مرحله در پیاده سازی مدل آریما در نرم افزار متلب جمع­آوری داده­ها می­باشد، در پروژه حاضر مجموعه دادهFoursquare استفاده شده است در مجموعه مقادیر، دادهایمان به صورت اطلاعات مهم و کاربردی از کاربران شبکه اجتماعی Foursquare می­باشد.

* User ID (anonymized)

درمجموعه داده­های پژوهش حاضر هر رکود کاربرآن دارای شناسه منحصر به فرد است که آنرا با سرستون آی دی[[11]](#footnote-11) در فایل اکسل مشخص می­کند.

* Venue ID (Foursquare)

در مجموعه داده­های پژوهش حاضر هرمکان آن دارای شناسه منحصر به فرد است که آنرا با سرستون آی دی در فایل اکسل مشخص می­کند.

* Venue category ID (Foursquare)

در مجموعه داده­های پژوهش حاضر طبقه بندی بر اساس آی دی مکان­ها در فایل اکسل قرار داده است.

* Venue category name (Fousquare)

در مجموعه داده­های پژوهش حاضراین فیلد براساس طبقه بندی بر اساس نام مکان­ها در فایل اکسل قرار داده شده است.

* Latitude

در مجموعه داده­های پژوهش حاضر این فلید بر اساس عرض جغرافیای در فایل اکسل قرار داده شده است.

* Longitude

در مجموعه داده­های پژوهش حاضر این فلید بر اساس طول جغرافیای در فایل اکسل قرار داده شده است.

* Timezone offset in minutes

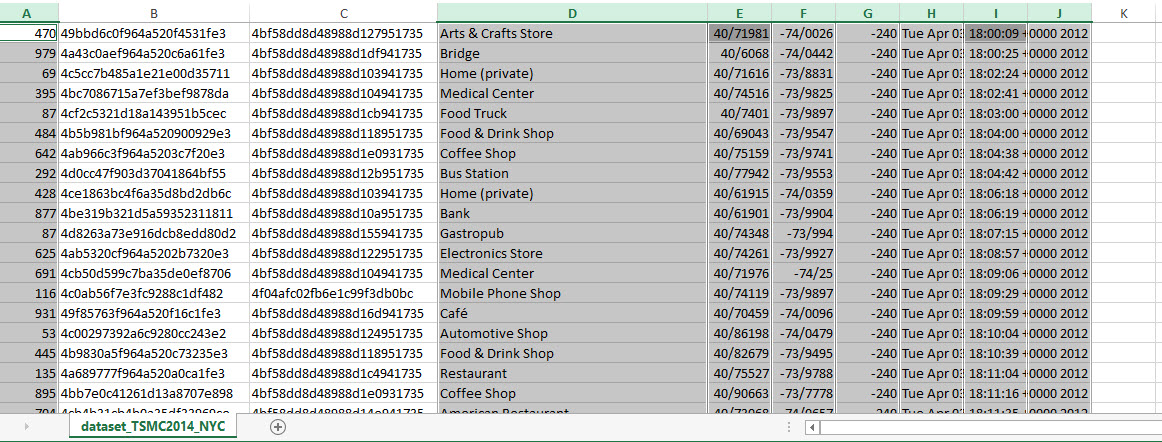
در مجموعه داده­های پژوهش حاضر این فلید بر اساس منطقه زمانی در دقیقه در فایل اکسل قرار داده شده است.

* UTC time

درمجموعه داده­های پروژه حاضر این فلید بر اساس زمان یو تی سی(ساعت هماهنگ جهانی) در فایل اکسل قرار داده است.

مجموعه داده ­های پروژه دارای1000 رکورد می باشد، که از12 آپریل سال 2012 تا 16فبریه سال 2013 از شهر نیوریک جمع­آوری شده است. مجموعه داده­های پژوهش حاضر برای مدل سازی به روش آریما برای توصیه مکان در شبکه اجتماعی مبتنی بر مکان در سری­های از زمان را مورد بررسی قرار خواهیم داد.

برای مدل سازی مدل آریما کافیست بر روی ستون شناسه (ستونA فایل اکسل )می­باشد. و ستون زمان(ستون I فایل اکسل می­باشد) را از مجموعه داده­های پژوهش حاضر برای برنامه نویسی از آن استفاده خواهیم کرد.



## شکل4-2 مجموعه داده های پروژه

روشي که برای مقایسه با روش پيشنهادی انتخاب شده است به شرح زیر مي باشد:

روش بن شیا[[12]](#footnote-12)

رویکرد پیشنهادی این مقاله در داده­ها چک می­کند براساس ثبت ترجیحات کاربران در مقوله زمانی-مکانی و اهمیت توصیه مکان(رتبه بندی[[13]](#footnote-13)) از یک مدل یادگیری ماشین استفاده می­کند در این تحقيق توصیه­های مکان های موجود مبتنی بر بافت است، اطلاعات زمان را با رکوردهای مراجعه کاربران به منظور جمع آوری ترجیحات کاربر (علاقه کاربر)ترکیب می کند و سپس دسته بندی می­کند تا به طور موثر دقت توصیه را بهبود می بخشد.استفاده از داده های "مراجعه" در جذب ترجیحات کاربران و استفاده از یک مدل یادگیری ماشین برای تنظیم اهمیت عوامل مختلف در رتبه بندی را بررسی می کند و از میان مدل های یادگیری ماشین روش های مانند SVM و RSVM و EESVM برای مدل سازی یادگیری ماشین هستند، روش ESSVM برای مدل سازی توصیه مکان در شبکه های اجتماعی مبتنی بر مکان از دیتاست FourSquare استفاده کرده است در این مقاله بر روی دو دیتا ست مشهور شبکه های اجتماعی مبتنی بر مکان مانند رسانه تویتر و FourSquare که تویتر از مشهور ترین شبکه های اجتماعی تحت وب می­باشد استفاده کرده است. در آدرس سایت ( https://dev.twitter.com/ ) مجموعه تویت­های عمومی در آن وجود دارد که می­توان از طریق فیلتر کردن کلمات به داده های مورد نظر برسیم در این مقاله از بر روی داده های FourSquare تولید شده اند فیلتر را بر اساس FourSquare انجام داده است که مجموعه داده بدست آمده ۴۱۹۵۰۹ تویت که از ۴۹۸۲۳ کاربر بدست آمده علاوه براین کلیه جزیئات از مکان های مورد بازیدید و مورد علاقه کاربران بوده را جمع آوری کرده است که سایت https://developer.foursquare.com/. جمع آوری شده است.

دلایل استفاده از روش ESSVM در این مقاله به شرح زیر می­باشد**.**

1. توصیه های مکان اغلب محبوب ترین، ارزان ترین یا نزدیک ترین مراکز را توصیه می کنند و این روش ها ترجیحات کاربران ناهمگون را در نظر نمی گیرد. برخلاف روش­های گذشته، توصیه های مکان مبتنی بر بافت، اطلاعات زمان و دسته بندی را با رکوردهای مراجعه کاربران به منظور جمع آوری ترجیحات کاربر ترکیب می کند که به طور موثر دقت توصیه را بهبود می بخشد.
2. ایده اصلی توصیه مکان مبتنی بر بافت ، رتبه بندی توصیه برای هر کاربر است. با این حال، رتبه بندی سنتی SVM، (RSVM) زمان زیادی را با افزایش تعداد مکان ها می برد، و دقت توصیه برجسته نیست. برای مقابله با این مسئله، بر روش SVM رتبه بندی مکان جاسازی شده (ESSVM) برای بهینه سازی عملکرد تکیه می شود و هم کارایی و هم تاثیر گذاری در عمل توصیه بهبود می یابند.

مراحل روش این مقاله به شرح زیر است.

 پيش پردازش داده

 مراحل آماده سازی داده: در این مقاله از مجموعه داده­های FourSquare استفاده شده است.

 مرحله رتبه بندی بر اساس بازدید کاربران

 مرحله دسته بندی: این مرحله شامل دو مرحله یادگيری و آزمایشي مي باشد که در ادامه هرکدام از این مراحل توضيح داده مي­شود :

o مرحله یادگيری در این مرحله با استفاده از الگوریتم یادگیری ماشین EESVM اقدام به ایجاد مدل برای طبقه بندی نظرات مي کنيم

o مرحله آزمایش: دانشي که در مرحله یادگيری مدل مي شود مي بایست در مرحله تست مورد تحليل قرار بگيرد تا بتوان ارزش و صحت آن را تعيين نمود و در پي آن کارایي مدل را نيز ارزیابي کرد

**معیارهای ارزیابی**

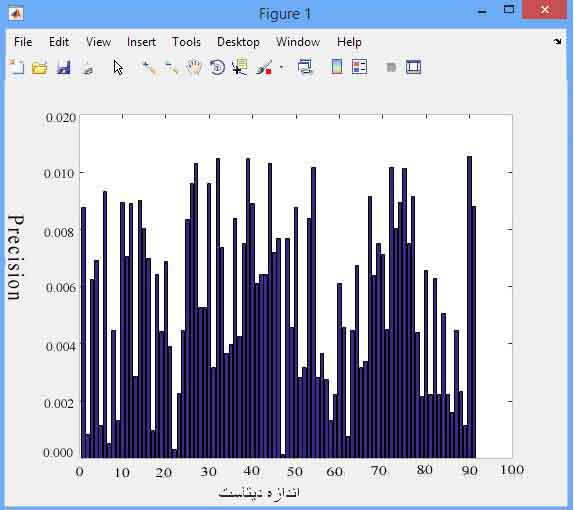
روش پيشنهادی با روش قبلي طبق چندین معيار مقایسه شده است کهه در ادامه هر یک توضيح داده

شده اند.

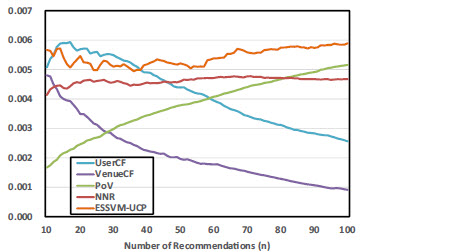
معیار ارزیابی دقتPrecision : ظرفیت یک سیستم توصیه گر برای نشان دادن فقط آیتم های)مکان ها( مفیدرا با تلاش برای به حداقل رساندن مخلوطی از آنها با آیتم های بی فایده نشان می دهد.

در اينجا TP ، FP و به ترتيب نشان دهنده ي True Positive ، False Positive مي باشند .

همانگونه که در شکل و نمودار های ارزیابی دقت پژوهش حاضر و تحقیق بن شیا مشاهده می­شود درهردو تحقیق بر روی ۱۰۰ رکورد ارزیابی انجام شده است و با توجه به نمودارهای بدست آمده دقت روش پژوهش حاضر بالاتر می­باشد.



نمودار دقت آریما



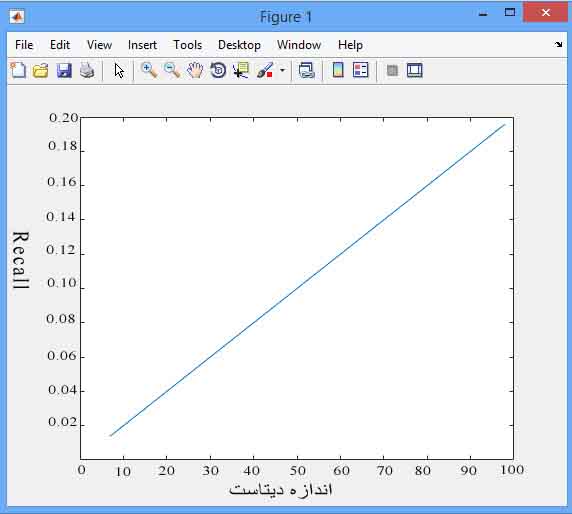
نمودار دقت بن شیا

معیارحساسیت : Recall نشان دهنده پوشش اقلام)مکان ها( مفیدی است که سیستم توصیه گر می

تواند بدست آورد. به عبارت دیگر ، این معیار اندازه گیری ظرفیت به دست آوردن تمام موارد مفید موجود در میان انبوه آیتم ها می باشد .

در اينجا TP و FN به ترتيب نشان دهنده ي True Positive و False Negative مي باشند.

* با مقایسه نمودار ها در شکل 4-4-۳-1 و 4-4-۳-2 به این نتیجه خواهیم رسید که مدل سازی با روش آریما دقت بهتری نسبت به مقاله بن شیا که مدل یادگیری ماشین است نشان داده می­دهد.



نمودار حساسیت مدل آریما

ایجاد ماتریس شناسه[[14]](#footnote-14)

همانگونه که در شکل 4-۲ از پژوهش حاضر دیده می­شود تعداد سطر و ستون های پژوهش زیاد می­باشد و همانگونه که در داده کاوی و کار بر روی پایگاه داده­ها مشهود هست ما باید فیلدهای خاص را مورد بررسی قرار بدهیم برای این کار ستون مربوط به شناسه رکودها که اعداد منحصر به فردی نیز هستند را در ماتریس جداگانه قرار داده ایم.

با اجرای دستور “نام ستون.نام جدول=نام ماتریس “ در محیط برنامه نویسی متلب ستون مربوط به شناسه را در ماتریسی قرار خواهیم داد که با شناسه هر رکورد اشاره ای از مکان برای هر سری زمانی خواهیم داشت.

## پس از آن به ترسیم داده های شناسه میپردازیم با دستور figure و دستور plot برای ترسیم داده های شناسه خواهیم پرداخت.

-ایحاد ماتریس زمان در محیط برنامه نویسی متلب

داده­های این ماتریس برای کار روی سری های زمانی است و همراه با ماتریس شناسه برای توصیه پر بازدید ترین مکان­ها در سری های زمانی مختلف از شبانه روز را خواهیم داشت.

روند برنامه نویسی آن مانند بخش قبل ایجاد ماتریس شناسه هست با این تفاوت که در ماتریس شناسه با ستون اول از مجموعه داده­های پژوهش استفاده کرده­ایم و برای این بخش از پژوهش از ستون های پنجم و ششم استفاده خواهیم کرد.برای اینکه حجم داده های پژوهش زیاد می‌باشد ما ۱۰0 رکورد را بررسی کرده ایم که در سری زمانی 12 ظهر تا ۳ بعد ازظهر کدوم مکان ها توصیه می­شوند را در شکل زیر قابل مشاهده است تفسیر خروجی به شرح زیر است و به ترتیب زیر به رکوردهای از ترجیحات کاربران در مکان ها اشاره دارد.

سورمه ای تیره:رکوردهای اداره جات اشاره دارد

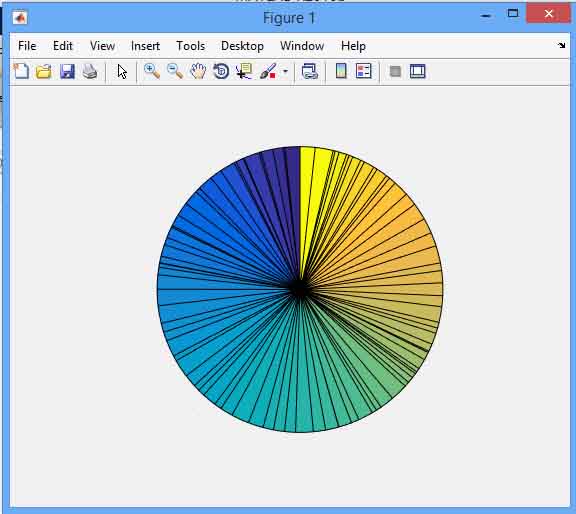
رنگ آبی تیره: به رکوردهای دانشگاه های بزرگ دارد

رنگ آبی کم رنگ : رکوردهای دانشگاه و مراکز آموزشی کوچک اشاره دارد

رنگ سبز: رکوردهای بانک اشاره دارد.

رنگ نارنجی: به رکوردهای رستوران ها و غذا فروشی و نوشیدنی فروشی اشاره دارد.

رنگ زرد: به رکوردهای مراکز تجاری اشاره دارد.



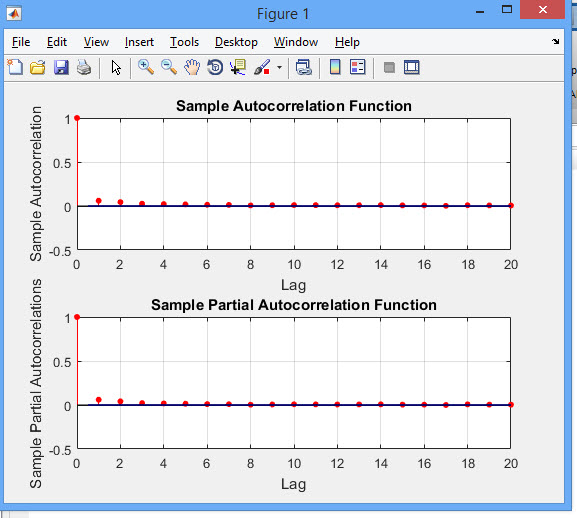
## نمودار سری زمانی

داده های پژوهش حاضر به درستی وارد نرم افزار متلب نموده ایم،از آنها خروجی­های به صورت نموداری و ماتریسی گرفته­ایم پس در نتیجه داده­ها به درستی قابل اجرا هستند.

نمودار خود همبستگی و خود هم بستگی جزئی

که برای بررسی اینکه آیا داده­های پروژه ایستا هستند از تابع و نمودار ها برای تشخیص ایستا بودن داده ها استفاده می­کنیم زیرا داده­های مدل آریما همانگونه که در فصل گذشته مفصل از آن گفته شد،بایدایستا باشند.

و در نهایت خروجی به شکل نمودار زیر مشخص می شود داده های که ما روی آنها پژوهش می­کنیم کاملا ایستا هستند.



ایجاد نمودار خود همبستگی. خود هم­بستگی جزئی

و نیز هم چنین می­توان از تابع زیر در محیط کدنویسی متلب بررسی نمود که آیا داده ها ایستا هستند و یا داده ها ایستا نیستند هر عددی این تابع به جز عدد یک به ما بدهد یعنی داده­های پژوهش ایستا نیستند.

pt = adftest(ID)

pt =

logical

1

خروجی تابع adftest پروژه ما یک شد پس داده­های پروژه حاضر ایستا هستند.

% diff

ایجاد ماتریس داده = diff(DATA);

figure,

مدل سازی آریما[[15]](#footnote-15)

پس از آنکه داده­های پروژه را بررسی کردیم و ایستایی آنرا به اثابت رسانده ایم. حال نوبت به مدل سازی مدل آریما خواهیم پرداخت برای اینکار ما از متد برنامه نویسی مدل سازی آریما استفاده خواهیم کرد کد نویسی آن به روش زیر است و در انتها از اجرای برنامه از متلب تصویری خواهیم گذاشت.

--------------------

Distribution: Name = 'Gaussian'

P: 1

D: 1

Q: 0

Constant: NaN

AR: {}

SAR: {}

MA: {}

SMA: {}

Variance: NaN

--------------------

Conditional Probability Distribution: Gaussian

Standard t

Parameter Value Error Statistic

----------- ----------- ------------ -----------

Constant 0.00220881 0.877871 0.0025161

Variance 175266 583.002 300.626

mdl =

ARIMA(0,1,0) Model:

--------------------

Distribution: Name = 'Gaussian'

P: 1

D: 1

Q: 0

Constant: 0.00220881

AR: {}

SAR: {}

MA: {}

SMA: {}

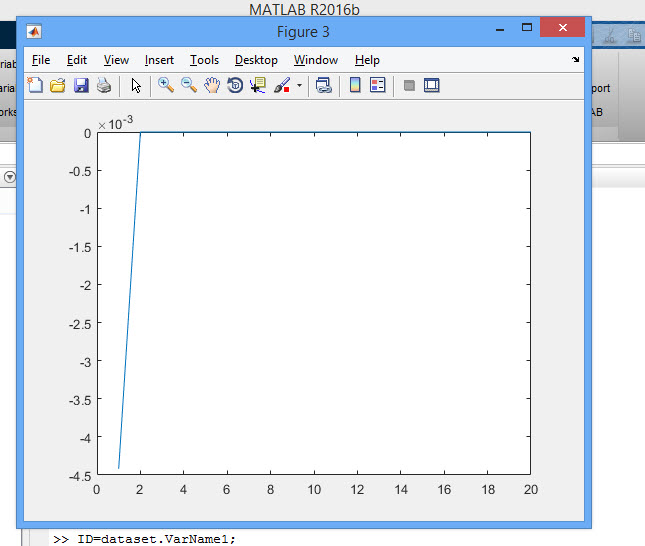
Variance: 175266

همانگونه که در تصویر بالا مشاهده می­کنید مدل آریما پروژه به درستی در محیط کاری متلب قرار گرفته و برنامه بدون خطا اجرا شده است.­

مدل ما با میانگین 0.00220881 و واریانس 175266 آموزش دیده و به درستی به اجرا رسیده است.

خروجی از میانگین مربعات خطا برای پیش­­بینی خطا نیز گرفته ایم در تصاویر زیر مشاهده می­شود.

خروجی پیش بینی ما با میانگین خطای صفر به درستی به اجرا رسیده است.

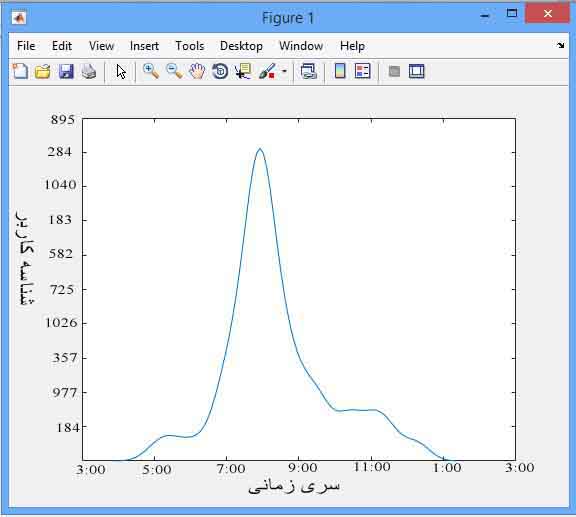


نمودارمیانگین مربعات خطا مدل آریما

داده كاوي،‎ ‎پايگاه­هاي‎ ‎داده‌هاي‎ ‎بزرگ را‎ ‎به عنوان‎ ‎منبع‎ ‎دانش‎ ‎در‎ ‎نظر‎ ‎مي­گيرد‎. فناوري­هاي نوين اطلاعاتي و ارتباطي، و همچنين تكنولوژي­هاي پشتيبان تصميم، با جمع­آوري، ذخيره، ارزيابي، تفسير و تحليل، بازيابي و اشاعه اطلاعات و دانش به كاربران خاص، مي­توانند در اطلاع­يابي به­موقع، صحيح و مورد­نياز به افراد تاثير زيادي داشته­باشند. در پژوهش حاضر از داده های خام مسئله با استفاده از تکنیک مدل سازی آریما به تفسیری از نتایج شبیه سازی رسیده ایم.در تحلیل سری­های زمانی به این نتیجه رسیده ایم که نتایج داده­ها نشان می­دهند که انتخاب مکان­ها در دوره­های زمانی مختلف شیوه زندگی و رفتار کاربران یک کشور، شهر یا منطقه جغرافیایی را نشان می­دهد.

زمان مهم ترین عامل انتخاب مکان در شبکه های اجتماعی مبتنی بر مکان می­باشد. در پژوهش حاضر به رفتارکاربران در طول روز و شب پرداخته شده است. ونشان داده می­شود که در طول روز کاربران مکان ها را انتخاب می­کنند که معمولا محیط کاری و آموزشی است و در طول شب به مکان­های تفریحی مانند انواع رستوران ها و بارها و کافه ها را انتخاب می­کنند. و این موارد تنها با بررسی سری­های زمانی حضور کاربران در شبکه­های اجتماعی مبتنی بر مکان میسر است.

در شکل زیر نتایج نشان می­دهد که­ ترجیحات افراد در این ساعت ها رفتن به رستوران ها و بارها می باشد بعنوان مثال کاربر شماره ۲۸۴ مکان بار را ترجیح داده است. و نیز کاربر با شماره آی دی[[16]](#footnote-16) 1040 نیز مکان بار را انتخاب کرده هر دو کاربر با شماره آی دی 284 و1040 در سری زمانی بین ساعت ۷ تا ۹ شب بوده و کاربر با شماره آی دی 1026 مکان رستوران مکزیکی را ترجیح داده است.



توزیع از هر دوره ی زمانی با توجه به دسته بندی های مختلف

1. trend [↑](#footnote-ref-1)
2. Cyclical variation [↑](#footnote-ref-2)
3. Seasonal varition [↑](#footnote-ref-3)
4. Irregular variation [↑](#footnote-ref-4)
5. Autoregressive [↑](#footnote-ref-5)
6. Thomas Fiering [↑](#footnote-ref-6)
7. Moving Average [↑](#footnote-ref-7)
8. Autoregressive Moving Average [↑](#footnote-ref-8)
9. ARMA [↑](#footnote-ref-9)
10. ARIMA [↑](#footnote-ref-10)
11. ID [↑](#footnote-ref-11)
12. Bin Xia [↑](#footnote-ref-12)
13. Ranking [↑](#footnote-ref-13)
14. ID [↑](#footnote-ref-14)
15. ARIMA [↑](#footnote-ref-15)
16. ID [↑](#footnote-ref-16)