

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

گروه مهندسی فناوری اطلاعات

پایان نامه کارشناسی

رشته مهندسی کامپیوتر گرایش فناوری اطلاعات

عنوان پروژه:

تحلیل زمانی رفتار ترافیکی شبکه با استفاده از الگوریتم‌های تحلیل سری زمانی

استاد راهنما:

دکتر بهروز شاهقلی

پژوهشگران:

علی هداوند

رضا پازن

شهریور ۱۴۰۰



دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

گروه مهندسی فناوری اطلاعات

پروژه کارشناسی رشته‌ی مهندسی کامپیوتر گرایش فناوری اطلاعات

آقایان علی هداوند و رضا پازن

تحت عنوان

تحلیل زمانی رفتار ترافیکی شبکه با استفاده از الگوریتم‌های تحلیل سری زمانی

در تاریخ / / ۱۳ توسط هیأت داوران زیر بررسی و با نمره به تصویب نهایی رسید.

۱- استاد راهنمای پروژه:

دکتر

امضا

۲- استاد داور :

دکتر

امضا

امضای مدیر گروه

تقدیم به

پدر و مادر عزیزمان، که تلاش بی دریغ شان روشنی بخش آینده ی ما
بود.

چکیده:

واژگان کلیدی: شبکه‌های کامپیوتری، تحلیل داده‌های جاری، سری زمانی، بازه‌های زمانی

فهرست مطالب

۱- فصل اول مقدمه.....	۱
۱-۱ بیان مسئله.....	۱
۲-۱ ارزش پروژه.....	۲
۳-۱ هدف پروژه.....	۲
۴-۱ رویکرد پیشنهادی.....	۳
۵-۱ ساختار پایان نامه.....	۳
۲- فصل دوم مفاهیم و کلیات.....	۴
۱-۲ مقدمه.....	۴
۲-۲ تحلیل جریان داده‌ی شبکه‌های کامپیوتری.....	۴
۳-۲ معرفی سری‌های زمانی.....	۶
۱-۳-۲ مدل‌های جمع‌آوری داده در سری زمانی.....	۶
۴-۲ مولفه‌های رفتاری سری زمانی.....	۷
۱-۴-۲ روند (Trend).....	۷
۲-۴-۲ فصلی‌بودن (Seasonality).....	۸
۳-۴-۲ الگوهای تناوب (Cyclic Patterns).....	۹
۴-۴-۲ خطاها (Errors).....	۹
۵-۴-۲ نمودارهای تابع خودهمبستگی (ACF) و خودهمبستگی جزئی (PACF).....	۱۰
۶-۴-۲ اختلال سفید (White Noise).....	۱۲
۷-۴-۲ پیاده‌روی تصادفی (Random Walk).....	۱۳
۸-۴-۲ ایستایی (Stationarity).....	۱۴
۵-۲ پیش‌بینی در سری زمانی (Forecasting).....	۱۷
۱-۵-۲ مدل‌های پیش‌گویی در سری‌های زمانی.....	۱۷
۶-۲ جمع‌بندی.....	۱۷
۳- فصل سوم روش پیشنهادی.....	۱۸
۱-۳ مقدمه.....	۱۸
۲-۳ جمع‌بندی.....	۱۸

۴- فصل چهارم نتایج.....	۱۹
۴-۱ مقدمه.....	۱۹
۴-۲ جمع بندی.....	۱۹
۵- فصل پنجم جمع بندی.....	۲۰
۶- پیوست.....	۲۱
۷- منابع.....	۲۳

فهرست شکل‌ها

- شکل ۱-۲: اطلاعات ذخیره شده در فایل pcap ۵
- شکل ۲-۲: نمونه‌ای از نمودار سری زمانی [۹] ۶
- شکل ۲-۳: نمونه‌ای از روند سری زمانی [۹] ۸
- شکل ۲-۴: خروجی نمودار تحلیل فصلی ۹
- شکل ۲-۵: نمونه‌ی داده‌های خطا در سری زمانی [۳] ۹
- شکل ۲-۶: تأخیرهای جمع‌آوری شده از چهار سری زمانی ۱۰
- شکل ۲-۷: نمودار ACF [۹] ۱۱
- شکل ۲-۸: نمودار PACF [۹] ۱۲
- شکل ۲-۹: نمونه‌ای از white noise [۹] ۱۲
- شکل ۲-۱۰: نمودار ACF برای سری زمانی White Noise [۹] ۱۳
- شکل ۲-۱۱: نمودار یک سری زمانی RW در کنار یک سری زمانی نرمال ۱۴
- شکل ۲-۱۲: نمودار ACF برای سری زمانی RW ۱۴
- شکل ۲-۱۳: نمودارهای white noise در تحلیل ایستایی ۱۵
- شکل ۲-۱۴: نتایج تست دیکی-فولر ۱۶
- شکل ۲-۱۵: نمودار ACF برای سری زمانی RW [۱۲] ۱۷
- شکل ۶-۱: صفحه‌ی ابتدایی Jupyter Notebook ۲۱
- شکل ۶-۲: سرور اجرا شده برای Jupyter Notebook ۲۱
- شکل ۶-۳: محیط یکپارچه‌سازی شده‌ی Jupyter در Visual Studio Code ۲۲
- شکل ۶-۴: صفحه‌ی ابتدایی Google Colab ۲۲

فهرست جدول‌ها

جدول ۱-۱: تست جدول اول.....**Error! Bookmark not defined.**

CN	Computer Networks
TS	Time Series
ACF	Auto Correlation Function
PACF	Partial Auto Correlation Function
AR	Auto Regressive
MA	Moving Average
ARMA	Auto Regressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
ETS	Error Trend Seasonality

فصل اول

مقدمه

۱-۱ بیان مسئله

در طول تاریخ یکی از اساسی‌ترین نیازهای انسان برقراری ارتباط بوده است که با گذر زمان، ابزار و روش‌های آن نیز دستخوش تغییر شده‌اند. امروزه، فناوری شبکه‌های کامپیوتری^۱ (CN) یکی از مهم‌ترین ابزارهای برآورده کردن نیازهای تعاملی انسان است. از ابتدای پیدایش این فناوری تا کنون، به دلیل گسترش جوامع و به دنبال آن گسترش استفاده از شبکه‌های کامپیوتری، میزان داده‌ی جاری در این شبکه‌ها افزایش یافته و چالش‌هایی نیز پیش روی کاربران و توسعه‌دهندگان این فناوری قرار داده است. مواردی مانند تامین محرمانگی، وجود یا عدم وجود خطا، صحت انتقال اطلاعات، تأخیر و غیره، از جمله این چالش‌ها هستند.

با توجه به گسترده شدن کاربرد شبکه‌های کامپیوتری و حضور این فناوری در تمام عرصه‌های زندگی انسان، از سازمان‌های بزرگ تا کاربری خانگی، حفظ سلامت این شبکه‌ها امر بسیار مهمی تلقی می‌شود. یکی از راه‌های کنترل این شبکه‌ها، پایش^۲ آن‌ها است. پایش به معنی جمع‌آوری و تحلیل بسته‌های^۳ انتقال داده شده در شبکه است [۱]. کارشناسان و متخصصان، با استفاده از ابزارها و روش‌های مختلف پایش شبکه مانند پروتکل مدیریت ساده‌ی شبکه^۴ (SNMP) و پایش بلادرنگ^۵ قادر به مشاهده، ذخیره و بررسی داده‌های شبکه هستند که به آن‌ها کمک می‌کند رفتار شبکه‌ی موردنظر را ثبت و در صورت بروز ناهنجاری^۶ یا خطا در جریان داده‌ها، آن را گزارش کنند [۱].

تحلیل داده‌های شبکه و تشخیص الگوی رفتاری آن کار آسانی نیست؛ زیرا جریان داده‌ای شبکه متغیری پیوسته در زمان است و فرایند تحلیل و نتیجه‌گیری در این مورد باید بازده قابل‌قبولی داشته باشد. یکی از اهداف تحلیل داده‌های شبکه را می‌توان ثبت رفتار آن دانست. در این راستا، در این پروژه با استفاده از مفاهیمی به نام سری‌های

¹ Computer Networks

² Monitoring

³ Packets

⁴ Simple Network Management Protocol

⁵ Real-Time Monitoring

⁶ Anomaly

زمانی^۱ (TS) در کنار علم آمار، روشی برای تحلیل داده‌های شبکه و دسته‌بندی^۲ رفتار آن در بازه‌های زمانی مختلف، جهت ثبت یک الگوی ثابت، ارائه شده است.

۲-۱ ارزش پروژه

در زمینه‌ی تحلیل داده‌های شبکه پروژه‌های مشابهی وجود دارد که خروجی آن‌ها با استفاده از ابزارهایی مانند زبان نشانه‌گذاری توسعه‌پذیر^۳ (XML) تولید می‌شود. در این پروژه‌ها، بازه‌های زمانی به صورت پیش‌فرض توسط کاربر تعریف می‌شوند و تحلیل جریان داده‌ی شبکه فقط در بازه‌های از پیش تعریف شده و به صورت ایستا انجام می‌شود. پروژه‌ها و مقاله‌های دیگر نیز که از مفاهیم سری زمانی برای تحلیل داده‌های شبکه استفاده کرده‌اند، به هدف تشخیص ناهنجاری و یا دسته‌بندی جریان داده‌ی شبکه بر اساس ماهیت داده‌ها بوده است؛ به طور مثال داده‌های شبکه‌های اجتماعی از داده‌های دیگر سرویس‌ها جداسازی شود.

در این پروژه، داده‌های موجود با استفاده از مفاهیم سری زمانی مورد بررسی قرار گرفته است و خروجی مورد نظر، بازه‌های زمانی هستند که هر بازه‌ی زمانی مشخص کننده‌ی قطعه‌ای از زمان است که میزان داده‌ی جاری در شبکه در آن زمان مقدار مشخصی دارد. این خروجی به سیستم و کاربران قالب مشخصی از رفتار شبکه را ارائه می‌دهد که در تصمیم‌گیری‌های آینده بسیار مؤثر است. همچنین قالب ورودی پروژه به صورت فایل است که این فایل شامل اطلاعات بسته‌های شناسایی شده‌ی جاری در شبکه است. این امر باعث افزایش قابلیت حمل برنامه‌ی پروژه می‌شود.

در نهایت با توجه به نیاز جدی نظارت بر رفتار و جریان داده‌های شبکه‌های کامپیوتری و همچنین پیچیدگی زیاد و زمان‌بر بودن استفاده از ابزارهای موجود برای تحلیل داده‌ها، پیاده‌سازی ابزاری که فرایند تحلیل و نتیجه‌گیری و در نهایت تصمیم‌گیری را به صورت خودکار ارائه می‌دهد و خروجی متفاوت و پرکاربردی تولید می‌کند امری قابل توجه است.

به طور کلی، با استفاده از مفاهیم سری زمانی می‌توان ویژگی‌های رفتاری داده‌های ورودی را مشخص کرد. استفاده از تحلیل سری زمانی، به کاهش خطاهای احتمالی پیش‌آمده در روند تحلیل داده‌ها کمک می‌کند و امکان پیش‌بینی^۴ رفتار آینده‌ی شبکه و همچنین پیاده‌سازی سیستم تشخیص ناهنجاری را نیز به کاربر می‌دهد.

۳-۱ هدف پروژه

هدف اصلی از انجام این پروژه، ارائه‌ی روشی جهت تحلیل جریان داده‌ی شبکه‌های کامپیوتری با استفاده از مفاهیم تحلیل سری‌های زمانی است. تحلیل‌ها و پردازش‌های انجام شده بر روی سری‌های زمانی داده‌های جاری در شبکه با هدف تولید بازه‌های زمانی است که الگوی رفتاری شبکه موردنظر را نتیجه می‌دهد.

به‌طور کلی این پروژه اهداف زیر را در نظر دارد.

¹ Time Series

² Classification

³ Extensible Markup Language

⁴ Forecasting

- افزایش بهره‌وری از داده‌های جاری در شبکه جهت شناسایی الگوی رفتاری
- ایجاد زمینه برای پیاده‌سازی سیستم‌های تشخیص ناهنجاری در شبکه
- دسته‌بندی رفتار شبکه طبق تغییرات جریان داده‌ها طی گذر زمان
- بهره‌گیری از تحلیل‌های آماری و سری زمانی در توصیف رویدادهای شبکه

۴-۱ رویکرد پیشنهادی

این پروژه با استفاده از داده‌های شبیه‌سازی شده انجام شده است. جهت تحلیل و پردازش‌های گسترده‌تر نیاز به داده‌های سالیانه از شبکه‌های فعال بود که دسترسی به این داده‌ها مشکل است.

به طور کلی مراحل انجام پروژه به صورت زیر است:

• مرحله‌ی اول: پردازش و آماده‌سازی داده‌ها

در این مرحله فایل داده‌های ورودی در برنامه بارگذاری شده و اطلاعات موردنیاز آن جداسازی می‌شود. همچنین قالب تاریخ و ساعت درج شده در فایل داده‌ها به قالب قابل پردازش در برنامه تبدیل شده و اندیس‌گذاری ستون‌ها بر اساس تاریخ و ساعت به دست آمده مرتب می‌شود.

• مرحله‌ی دوم: تحلیل داده‌های پردازش شده و استخراج ویژگی‌های اولیه

در این مرحله با استفاده از توابع آماده‌سازی شده در کتابخانه‌های اضافه شده به برنامه، نمودارهای داده‌های پردازش شده را رسم کرده و ویژگی‌های سری زمانی به دست آمده را بررسی و ثبت می‌کنیم.

• مرحله‌ی سوم: پردازش نهایی، تعریف پنجره‌های اولیه و دریافت خروجی

در این مرحله با تعریف پنجره‌های اولیه و محاسبه‌ی واریانس و میانگین داده‌ها، تحلیل نهایی انجام شده و خروجی مورد نظر تولید می‌شود.

۵-۱ ساختار پایان‌نامه

این پایان‌نامه شامل هفت فصل است. در فصل دوم مفاهیم تحلیل داده‌های شبکه‌های کامپیوتری، ابزارها، روش‌ها و در نهایت مفاهیم موردنیاز جهت تحلیل و بررسی نتایج اولیه‌ی سری زمانی در کنار بررسی تحقیقات انجام شده بیان شده‌اند. در فصل سوم رویکرد پیشنهادی به طول کامل بررسی شده و مسئله‌ی بیان شده به تفصیل بررسی می‌شود. در فصل چهارم نتایج به دست آمده ارائه شده و به تحلیل و بررسی آن‌ها به همراه ارائه‌ی کدهای مهم و ابزارهای استفاده شده پرداخته شده است. در فصل پنجم جمع‌بندی و نتیجه‌گیری نهایی بیان شده است. در نهایت بخش ششم پیوست‌ها و بخش هفتم منابع را ارائه می‌دهند.

فصل دوم

مفاهیم و کلیات

۱-۲ مقدمه

با توجه به گسترش علوم و فنون تحلیل داده‌ها، روش‌های زیادی در زمینه‌های تخصصی مختلف ارائه شده است. در این مورد، تحلیل سری‌های زمانی در دهه‌های اخیر بسیاری از محققان را به خود جذب کرده است. این پروژه با استفاده از مفاهیم موجود در علم تحلیل سری‌های زمانی، داده‌های جمع‌آوری شده از شبکه‌های کامپیوتری را مورد بررسی قرار می‌دهد.

روش‌های زیادی در زمینه‌ی تحلیل جریان داده‌ی شبکه و با اهداف مختلف ارائه شده است که در این فصل در کنار مرور این روش‌ها، به بررسی تعاریف و مفاهیم ابتدایی سری‌های زمانی، الگوریتم‌ها و مدل‌های موجود و نکات آن پرداخته شده است. همچنین تحقیقات انجام شده در این زمینه نیز مورد بررسی قرار گرفته و تعاریف بیان شده در آن‌ها توضیح داده شده است.

۲-۲ تحلیل جریان داده‌ی شبکه‌های کامپیوتری

امروزه، اهداف مدیران شبکه‌های کامپیوتری از تحلیل داده‌های شبکه به نگهداری وضعیت کنونی شبکه خلاصه نمی‌شود. با پیشرفت علوم هوش مصنوعی و داده‌پردازی، این امکان برای مدیران شبکه فراهم شده است تا با به‌کارگیری این فناوری‌ها در کنار مفاهیم شبکه‌های کامپیوتری، رفتار آینده‌ی شبکه‌ها را پیش‌بینی کرده، داده‌های جاری در آن‌ها را دسته‌بندی و ناهنجاری‌های رخ داده در شبکه را شناسایی و یا حتی پیش‌بینی کنند. همچنین با استفاده از ابزارهای پایش موجود می‌توان داده‌های شبکه را به صورت بلادرنگ جمع‌آوری و در قالب گزارشات در فرمت‌های مرسوم pcap و یا csv ذخیره کرد.

از آنجایی که امروز یافتن، ثبت و پیش‌بینی الگوی رفتاری شبکه‌های کامپیوتری اهمیت زیادی دارد، تکنیک‌هایی از علوم مختلف در این زمینه استفاده شده است که هرکدام اهداف و ویژگی‌های مشخصی دارد. به طور مثال، حافظه‌های طولانی کوتاه‌مدت^۱ (LSTM)، که یک معماری در علم یادگیری عمیق^۲ (DL) است، در کنار زمینه‌های مختلف شبکه،

^۱ Long Short-Term Memory

^۲ Deep Learning

مانند اینترنت اشیا^۱ (IoT) با اهداف مختلفی مانند پیش‌بینی داده‌های شبکه پیاده‌سازی می‌شود [۱۳، ۱۴]. همچنین تحلیل‌های آماری نیز در مورد داده‌های شبکه کاربرد زیادی دارند. به طور مثال، مفاهیم سری زمانی، بیشتر از جنبه‌های آماری به تحلیل داده‌های شبکه می‌پردازند. استفاده از مدل‌های مختلف این الگوریتم‌ها به طور ترکیبی نیز مرسوم است. استفاده از شبکه‌های LSTM در کنار سری‌های زمانی جهت پیش‌بینی و یا دسته‌بندی کردن داده‌ها نمونه‌ای از این کاربردها است.

اولین مرحله برای شروع فرایند تحلیل داده‌ها، جمع‌آوری آن‌هاست. ابزارهای زیادی برای پایش و جمع‌آوری داده‌های شبکه وجود دارد که به صورت بلادرنگ به جمع‌آوری داده‌های جاری می‌پردازند. نرم‌افزارهایی مانند وایرشارک و یا اتریل ابزارهای پایش شبکه هستند که امکان جمع‌آوری و ذخیره‌سازی داده‌ها را در فایل‌هایی با فرمت pcap به کاربر می‌دهند. فایل‌های حاصل را می‌توان پایگاه‌داده‌ای از بسته‌های جاری در شبکه دانست که حاوی اطلاعاتی مانند آدرس مبدا، آدرس مقصد، حجم بسته و غیره هستند.

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	172.16.133.57	68.64.21.62	UDP	1168	53807 → 1853 Len=1126
2	0.000050	172.16.133.57	68.64.21.62	UDP	1168	53807 → 1853 Len=1126
3	0.000050	172.16.133.57	68.64.21.62	ADwin ...	94	
4	0.000322	96.43.146.176	172.16.133.82	TCP	60	443 → 61228 [ACK] Seq=1 Ack=1 Win=9659 Len=0
5	0.001160	172.16.133.56	68.64.21.42	UDP	167	49514 → 1853 Len=125
6	0.001306	68.64.21.62	172.16.133.57	UDP	67	1853 → 53807 Len=25
7	0.001307	96.43.146.176	172.16.133.82	TCP	60	443 → 61228 [ACK] Seq=1 Ack=10765 Win=0 Len=0
8	0.005263	96.43.146.176	172.16.133.82	TCP	60	443 → 60073 [ACK] Seq=1 Ack=1 Win=65535 Len=0
9	0.005988	172.16.133.49	68.64.21.41	UDP	167	58246 → 1853 Len=125

شکل ۱-۲: اطلاعات ذخیره شده در فایل pcap

شکل ۱-۲ داده‌های موجود در یک فایل pcap، که توسط نرم‌افزار wireshark تهیه شده است را نشان می‌دهد. لازم به ذکر است که در این پروژه، فایل‌های pcap به عنوان ورودی مورد استفاده قرار خواهند گرفت.

در کنار نرم‌افزارهای موجود، ابزارهای جامعی مانند پروتکل‌های^۲ مدیریت و نظارت بر شبکه‌های کامپیوتری ارائه شده است که SNMP یکی از این موارد است. پروتکل‌های پایش شبکه از راه دور^۳ (RMON) و پروتکل اطلاعاتی مدیریت مشترک^۴ (CMIP) موارد دیگر این ابزارها هستند [۲].

در این پروژه از مفاهیم سری زمانی استفاده شده است که در بخش بعدی به طور کامل توضیح داده خواهد شد. ابزارهایی تحت این مفاهیم نیز وجود دارد که به کاربر امکان جمع‌آوری و ذخیره‌سازی داده‌های جاری، نه تنها در شبکه‌های کامپیوتری، بلکه در تمامی اجزای سیستم را می‌دهد. به طور مثال، کاربر می‌تواند داده‌های مربوط به پردازنده‌ی سیستم خود را با استفاده از نمودارهای سری زمانی مشاهده کند. ابزارهایی مانند پرومتهوس^۵ از این دسته هستند که از پایگاه داده‌های سری زمانی جهت ذخیره‌ی داده‌ی خود استفاده می‌کنند. ابزارهایی نیز مانند گرافانا^۶ وجود دارند که به مصور سازی هرچه بهتر داده‌ها با نمودارها کمک می‌کنند. همچنین پایگاه داده‌های سری زمانی

¹ Internet of Things

² Protocols

³ Remote Network Monitoring

⁴ Common Management Information Protocol

⁵ Prometheus

⁶ Grafana

مانند اینفلاکس دی بی^۱ توسعه داده شده‌اند تا در صورت بهره‌گیری از الگوریتم‌های سری زمانی، بتوان داده‌ها را در آن‌ها ذخیره کرد.

۲-۳ معرفی سری‌های زمانی

به طور کلی مقادیری از داده که وابسته به زمان هستند و می‌توان طی گذر زمان آن‌ها را جمع‌آوری کرد، تشکیل یک سری زمانی می‌دهند. داده‌های سری زمانی می‌توانند متعلق به تغییرات دما، تغییرات جمعیت یک منطقه، تغییرات ارزش سهام در بازار بورس و یا داده‌های جمع‌آوری شده از یک CN باشند پس این داده‌ها، بیانگر تغییرات ایجاد شده در یک پدیده در طول زمان را منعکس می‌کنند.

به دلیل وابسته بودن داده‌های TS به زمان، می‌توان یک بردار مانند X در نظر گرفت و سری زمانی را به صورت زیر معرفی کرد [۳]:

$$X(t), t = 0, 1, 2, \dots$$

در این عبارت t بیانگر زمان و X یک متغیر تصادفی است. همان طور که در عبارت بیان شده است، زمان صفر نیز قابل استفاده است. این زمان، می‌تواند لحظه‌ی شروع یک پدیده و یا لحظه‌ی شروع جمع‌آوری داده‌های یک پدیده‌ی در جریان باشد [۳]. وابسته بودن داده‌های TS به زمان و تغییراتی که منعکس می‌کنند اهمیت ترتیب را در آن‌ها نشان می‌دهد. اگر در هر یک از مراحل جمع‌آوری، تحلیل و پیش‌بینی ترتیب داده‌ها دستخوش تغییر شود، نتایج به دست آمده قابلیت اعتماد ندارند.



شکل ۲-۲: نمونه‌ای از نمودار سری زمانی [۹]

شکل ۲-۱ نشان‌دهنده‌ی نمودار یک سری زمانی استخراج شده از داده‌های یک بازار سهام است. در نهایت می‌توان با استفاده از ابزارهای ترسیم مختلف نمودار، مقادیر داده‌های سری زمانی را به طور پیوسته در زمان ترسیم کرد.

۲-۳-۱ مدل‌های جمع‌آوری داده در سری زمانی

در جمع‌آوری داده‌های TS اگر فقط از یک ویژگی پدیده‌ی موردنظر استفاده شود، متغیر X در عبارت بیان شده یک‌بعدی بوده و مدل سری زمانی را یک متغیره^۲ می‌نامند. ولی اگر از چندین ویژگی برای جمع‌آوری داده استفاده

^۱ influxDB

^۲ Univariate

شود، به مدل سری زمانی چند متغیره^۱ گویند. وابستگی داده‌های TS به زمان امری اساسی است. ولی اگر در کنار ویژگی متغیر بودن با زمان تغییرات مکان و مختصات داده‌ها نیز لحاظ شود، مباحث مورد بحث وارد علم آمار فضایی^۲ می‌شوند.

معمولاً مرحله‌ی جمع‌آوری داده‌های TS بدون توقف و به صورت پیوسته انجام می‌شود که به آن زمان-پیوسته^۳ گویند؛ در غیر این صورت مدل جمع‌آوری داده را زمان-گسسته^۴ می‌نامند. از مثال‌های نام برده شده در بخش قبل، تغییرات جمعیت یک منطقه مثالی از مدل زمان-گسسته و تغییرات دما مثالی از مدل زمان-پیوسته هستند. در تحلیل‌های انجام شده معمولاً از روش جمع‌آوری زمان-گسسته انجام می‌شود که مقاطع مشخصی از زمان برای آن‌ها در نظر گرفته می‌شود. به طور مثال جمع‌آوری داده‌ها به صورت بازه‌های ساعتی، روزانه، هفتگی، ماهانه و سالانه انجام می‌شود. در نهایت سری‌های زمان-پیوسته قابلیت تبدیل شدن به سری‌های زمان-گسسته را دارند.

در این پروژه مدل جمع‌آوری داده‌های شبکه به صورت یک‌بعدی و زمان-گسسته است. همچنین داده‌های شبکه فقط از نظر تغییرات زمانی قابل تحلیل و بررسی هستند و مباحث علم آمار فضایی در این پایان‌نامه بررسی نمی‌شوند.

۲-۴ مولفه‌های رفتاری سری زمانی

داده‌های بررسی شده در سری زمانی، داده‌های پیوسته در زمان هستند که با گذر زمان رفتار متفاوتی از خود نشان می‌دهند. همیشه برای شروع تحلیل داده‌های سری زمانی، ابتدا باید ویژگی‌های مشخصی را از رفتار آن‌ها استخراج کرد تا بتوان با توجه به ویژگی‌های رفتاری داده‌ها بهترین مدل تحلیل را انتخاب و در نتیجه دقیق‌ترین پیش‌بینی را ارائه کرد.

در این پروژه، شناسایی این ویژگی‌ها کمک شایانی به استخراج خروجی موردنظر، یعنی بازه‌های زمانی می‌کند. این ویژگی‌ها به طور کلی نشان‌دهنده‌ی روند^۵ کلی داده‌ها به صورت صعودی یا نزولی طی گذر زمان، وجود تکرار در رفتار این روند و خطاهای احتمالی هستند که با استفاده از ابزارهای ترسیم نمودار، می‌توان آن‌ها را به صورت مصور نشان داد.

۲-۴-۱ روند (Trend)

اگر داده‌های یک سری زمانی در یک بازه‌ی مشخص از زمان به طور کلی صعودی و یا به طور کلی نزولی باشند، دارای روند مشخص هستند. روند را می‌توان با رسم نمودار سری زمانی تشخیص داد؛ به این صورت که در نمودار رسم شده، اگر نقطه‌ی ابتدایی و انتهایی نمودار سری زمانی به هم وصل شوند، شیب خط به‌دست‌آمده نشان‌دهنده‌ی روند کلی داده‌ها است. شکل ۲-۲ نمونه‌ای از نمودار سری زمانی را به همراه روال کلی آن که با خط قرمز نشان داده شده، نشان می‌دهد.

¹ Multivariate

² Spacial Statistics

³ Continuous Time

⁴ Discrete Time

⁵ Trend



شکل ۳-۲: نمونه‌ای از روند سری زمانی [۹]

باید به این نکته توجه داشت که روند دائمی نیست، و هنگامی که افزایش یا کاهش ممتد در داده‌ها رویت می‌شود می‌توان وجود روند را اعلام کرد. به طور کلی می‌توان گفت روند، تغییرات بلندمدت مقادیر سری زمانی را نشان می‌دهد. اگر داده‌ها در یک بازه‌ی زمانی مشخص تغییرات افزایشی و کاهشی نداشته باشند روند وجود ندارد.

علاوه بر ترسیم نمودار زمانی، توابع دیگری در تحلیل سری زمانی وجود دارند که نمودار روند را به طور جداگانه ترسیم می‌کنند. در ادامه به توضیح این توابع خواهیم پرداخت.

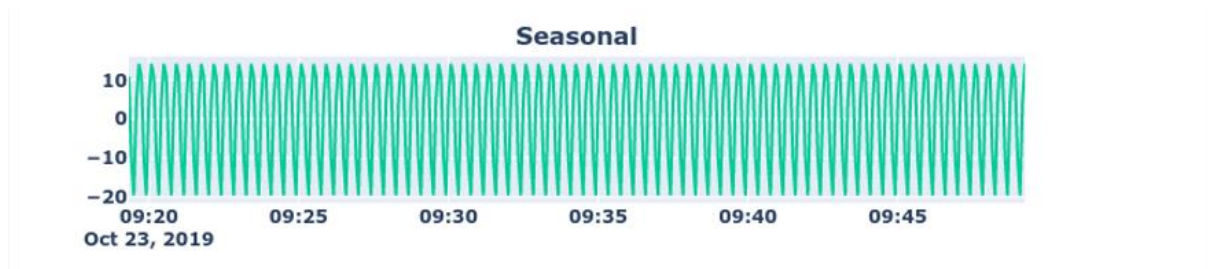
۲-۴-۲ فصلی بودن (Seasonality)

در مدت زمان یک سال، ۴ فصل با ویژگی‌های منحصر به فرد وجود دارد. زمان شروع و پایان هر فصل از پیش مشخص و تعیین شده است و با فرارسیدن هر کدام، تغییراتی در روند چرخش زمین و خورشید، دما و چرخه‌ی طبیعت دیده می‌شود که به طور کلی این تغییرات در هر سال یکنواخت و مشخص هستند. هیچ‌گاه دیده نمی‌شود که در اواسط سال شاهد سرد و زمستانی شدن هوا باشیم.

ویژگی فصلی بودن در داده‌های سری زمانی، مانند فصول سال، موعد شروع و پایان مشخصی دارد و فاصله‌ی بین هر رخداد آن ثابت و مشخص است. در طول یک فصل مشخص، تغییرات یکسانی بر داده‌های سری زمانی اعمال می‌شود. به طور مثال، اگر ویژگی فصلی بودن داده‌ها را به صورت سالیانه در نظر بگیریم و تغییرات سه ماه ابتدایی سال دارای خاصیت فصلی باشند، در سه ماه ابتدایی سال بعد نیز شاهد همان تغییرات خواهیم بود.

تغییرات فصلی محدود به سال نیستند. فصول تعریف شده می‌توانند محدود به ماه، هفته، روز و حتی ساعت باشند. به عنوان مثال اگر سری زمانی تغییرات میزان خرید کارمندان را در نظر بگیریم، با شروع هر ماه کارمندان حقوق خود را دریافت می‌کنند و ۱۰ روز ابتدایی هر ماه به انجام خریدهای مشخصی مشغول هستند. پس تغییرات مشخصی در ۱۰ روز ابتدایی هر ماه بر داده‌ها اعمال شده است.

نمودار خاصیت فصلی بودن را با استفاده از تابع مخصوصی که در تحلیل سری زمانی ارائه شده می‌توان استخراج کرد که خروجی آن شامل نمودار روند، نمودار فصلی و باقی‌مانده‌هاست که در بخش بعد توضیح داده شده است. شکل ۳-۲ نمونه‌ای از خروجی این تابع را برای قسمت فصلی نشان می‌دهد.



شکل ۴-۲: خروجی نمودار تحلیل فصلی

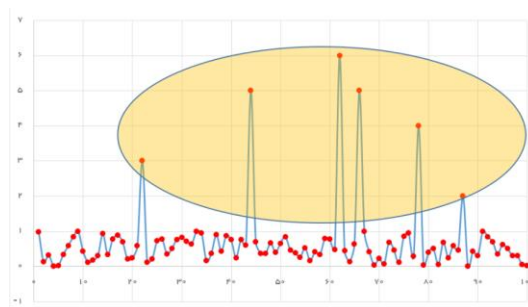
۳-۴-۲ الگوهای تناوب (Cyclic Patterns)

تغییرات چرخه‌ای که در بازه‌های مختلف و نامعلوم در روند داده‌ها دیده می‌شود، الگوهای تناوبی نام دارند. برخلاف الگوهای فصلی که در زمان مشخص و در بازه‌های تعیین شده امکان حضور داشتند، تغییرات تناوبی در بازه‌های نامشخص و طولانی مدت رخ می‌دهند که معمولاً این بازه‌ها بیشتر از ۲ سال است. به عنوان مثال چرخه‌ی ۴ مرحله‌ای کسب‌وکار که در طی ۳ سال رخ می‌دهد، باعث تکرار شدن روند داده‌های یک سازمان می‌شود و دارای خاصیت تناوبی است [۳]. فاصله‌ی میان هر دو رخداد تناوبی نیز از قبل مشخص نبوده و متغیر است.

۴-۴-۲ خطاها (Errors)

در بعضی سری‌های زمانی ممکن است قسمتی از داده‌ها نه روند خاصی داشته باشند و نه دارای خاصیت تکرار شونده‌ای مانند تناوب و فصلی بودن باشند. در تحلیل ابتدایی سری زمانی به این قسمت از داده‌ها، داده‌های خطا گویند زیرا توسط هیچ الگویی شناسایی نمی‌شوند. در برخی منابع نیز از این نوع داده‌ها به نام تغییرات نامعمول^۱ نام برده شده است.

این داده‌ها، که معمولاً به صورت فراز شدید^۲ و یا سقوط شدید^۳ در نمودار سری زمانی قابل رؤیت هستند، باید در مراحل ابتدایی تحلیل شناسایی و حذف شوند. زیرا خطاها هم‌بستگی داده‌ها را این برده و انجام اعمال بیشتر مانند پیش‌بینی سری زمانی توسط این نوع داده‌ها دچار اختلال می‌شود و نتایج گمراه‌کننده‌ای را تولید می‌کند.



شکل ۴-۵: نمونه‌ی داده‌های خطا در سری زمانی [۳]

در شکل ۴-۵ مقادیری که در ناحیه‌ی زردرنگ وجود دارند داده‌های خطا محسوب می‌شوند.

^۱ Irregular Changes

^۲ Spike

^۳ DoWNfall

۲-۴-۵ نمودارهای تابع خودهمبستگی^۱ (ACF) و خودهمبستگی جزئی^۲ (PACF)

جهت افزایش سادگی در پروسه‌ی تحلیل سری زمانی مقادیری از سری به عنوان نماینده انتخاب می‌شوند که به آن‌ها تأخیر^۳ گویند. فاصله‌ی میان تأخیرهای انتخاب شده یکسان و مشخص است. به طور مثال می‌توان از مقدار صفر شروع کرده و با فاصله‌ی زمانی بسیار اندکی از سری زمانی نمونه‌برداری کرد. در شکل ۲-۵ چهار سری زمانی قرار دارند که به جز تأخیر شروع، باقی تأخیرها با فاصله‌ی یک روزه از یکدیگر استخراج شده‌اند. موضوعی که در تحلیل سری‌های زمانی اهمیت دارد، بررسی میزان تغییرات میان این تأخیرها است. اگر سری زمانی موردنظر با سری زمانی دیگری مقایسه شود به آن بررسی میزان همبستگی گویند و اگر مقادیر یک سری زمانی با خودش مقایسه شود به آن بررسی میزان خودهمبستگی گویند.

1	date	spx	dax	ftse	nikkei
2	7/1/1994	469.9	2224.95	3445.98	18124.01
3	10/1/1994	475.27	2225	3440.58	18443.44
4	11/1/1994	474.13	2228.1	3413.77	18485.25
5	12/1/1994	474.17	2182.06	3372.02	18793.88
6	13/01/1994	472.47	2142.37	3360.01	18577.26
7	14/01/1994	474.91	2151.05	3400.56	18973.7
8	17/01/1994	473.3	2115.56	3407.83	18725.37
9	18/01/1994	474.25	2130.35	3437.01	18514.55
10	19/01/1994	474.3	2132.52	3475.15	19039.4

شکل ۲-۶: تأخیرهای جمع‌آوری شده از چهار سری زمانی

خودهمبستگی یک سری زمانی به معنی وجود همبستگی میان مقدار هر تأخیر با مقادیر تأخیرهای قبلی همان سری زمانی است. به عنوان مثال اگر x_n مقدار یک تأخیر در سری زمانی باشد میزان همبستگی آن با مقدار x_{n-1} با یک ضریب عددی مشخص می‌شود که این ضریب می‌تواند مقادیر مثبت و منفی میان صفر و یک داشته باشد. اگر ضریب همبستگی بین دو تأخیر از سری زمانی برابر عدد یک باشد به این معنی است که تغییری در مقدار جدید حاصل نشده و میزان همبستگی صد درصد است. ولی اگر تغییرات روی داده‌ها اعمال شده باشد ضریب همبستگی کمتر از عدد یک است. جهت استخراج ضرایب خودهمبستگی در یک سری زمانی، ابتدا تأخیرها به دست می‌آیند سپس مقادیر تأخیرها یک به یک بررسی شده و پس از هر بررسی، به میزان یک واحد تأخیر سری زمانی جابه‌جا^۴ می‌شود تا اختلاف مقدار هر تأخیر با مقادیر دیگر به دست آید.

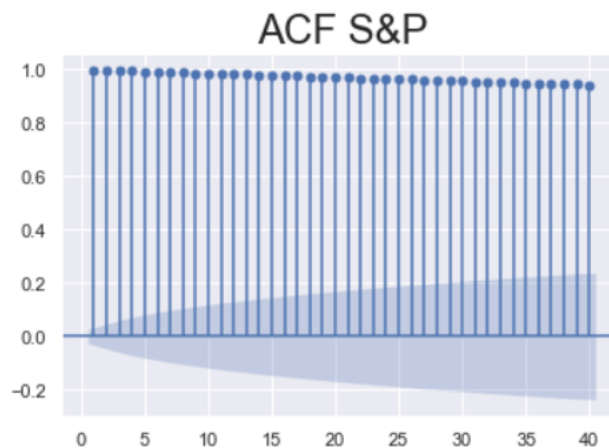
همان طور که از تعریف بیان شده برداشت می‌شود، برای تعیین میزان خودهمبستگی یک سری زمانی باید مقادیر تأخیرهای آن را مقادیر تأخیرهای قبلی در همان سری زمانی مقایسه کرد. تابع خودهمبستگی یا ACF سری زمانی به همراه تعداد تأخیرهای موردنیاز جهت بررسی را به عنوان ورودی دریافت کرده و نمودار ACF را رسم می‌کند.

¹ Auto Correlation Function

² Partial Auto Correlation Function

³ Lag

⁴ Shifts



شکل ۷-۲: نمودار ACF [۹]

ماهیت نمودار ACF به صورت نمودار میله‌ای است که هر میله‌ی آن نشان‌دهنده‌ی مقدار یک تأخیر سری زمانی است. در این نمودار، محور افقی نشان‌دهنده‌ی شماره و تعداد تأخیرها و محور عمودی نشان‌دهنده‌ی ضریب همبستگی است. همیشه ضریب همبستگی تأخیر صفر برابر یک است زیرا مقدارش با خودش مقایسه شده و به دلیل عدم وجود تفاوت همبستگی کامل وجود دارد. در نمودار ACF یک قسمت تیره روی محور افقی نمایش داده می‌شود که به آن فاصله‌ی اطمینان^۱ گویند. اگر ضریب همبستگی یک تأخیر در این ناحیه قرار گیرد، نشان‌دهنده‌ی ضریب همبستگی نزدیک به صفر است که بیان می‌کند همبستگی میان مقادیر آن تأخیر و مقادیر اطراف آن وجود ندارد. در نقطه‌ی مقابل، اگر مقدار ضریب همبستگی خارج از این محدوده باشد، میزان همبستگی مقادیر سری زمانی در آن تأخیر از نظر آماری قابل توجه است.

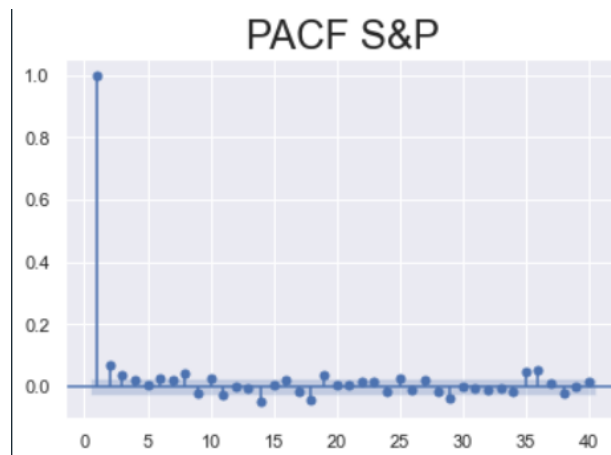
نمودار PACF که برای تحلیل‌های پیش‌بینی استفاده می‌شود، خروجی کاملاً مشابهی از نظر ظاهری با نمودار ACF دارد. یکی از تفاوت‌های نمودار ACF و PACF این است که در ACF مقادیر تأخیرها از خود سری زمانی برداشته می‌شوند ولی در PACF مقادیر تأخیرها از اختلاف مقادیر پیش‌بینی شده^۲ و مقادیر اصلی رؤیت شده^۳ در سری زمانی برداشت می‌شوند که به این اختلاف باقی‌مانده^۴ می‌گویند.

^۱ Confidence Interval

^۲ Forecasted Values – Fitted Values

^۳ Observed Values

^۴ Residual



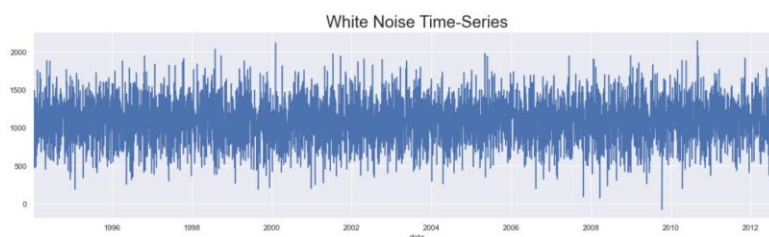
شکل ۸-۲: نمودار PACF [۹]

از کاربردهای نمودارهای معرفی شده می‌توان به مواردی مثل پیش‌بینی سری زمانی، تشخیص اختلال سفید^۱ (WN)، تعیین ایستایی^۲ سری زمانی و شناسایی پیاده‌روی تصادفی^۳ (RW) اشاره کرد که مفاهیم آن‌ها در ادامه بررسی شده‌اند.

۲-۴-۶ اختلال سفید (White Noise)

یکی از انواع سری‌های زمانی اختلال سفید است. میان مقادیر داده‌های این نوع سری زمانی هیچ‌گونه همبستگی وجود ندارد در نتیجه قابل پیش‌بینی نیست. داده‌های WN به زمان وابستگی نداشته و دارای میانگین صفر و واریانس ثابت (σ^2) هستند [۹، ۱۰]. در صورت تشخیص WN در یک سری زمانی، نمی‌توان تحلیل و بررسی بیشتری انجام داد. شکل ۲-۷ نمونه‌ای از سری زمانی WN را نشان می‌دهد. از دیگر ویژگی‌های WN می‌توان به عدم حضور روند در آن اشاره کرد. همچنین اگر در یک سری زمانی طولانی داده‌های خطا را جدا کرده و با آن‌ها سری زمانی جدیدی تشکیل دهیم، سری حاصل WN است.

راه‌های مختلفی جهت تشخیص WN بودن یک سری زمانی وجود دارد. به طور مثال می‌توان میانگین را محاسبه کرد که در صورت صفر بودن می‌توان گفت سری زمانی موردنظر WN است.



شکل ۹-۲: نمونه‌ای از white noise [۹]

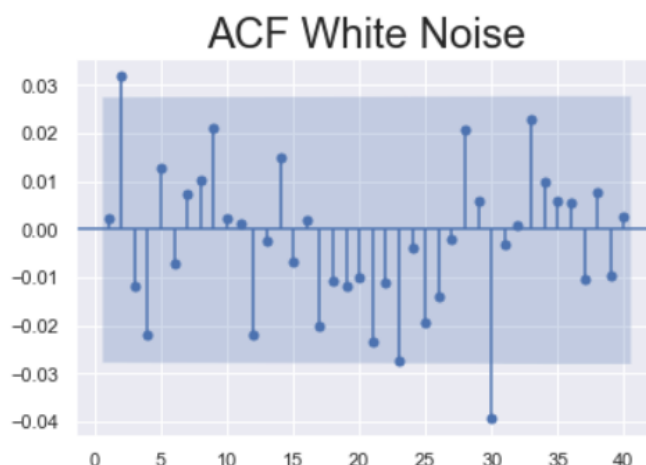
¹ White Noise

² Stationarity

³ Random Walk

همچنین در صورت محاسبه‌ی واریانس داده‌ها و بررسی تغییرات آن با گذشت زمان، در صورت ثابت بودن می‌توان ادعا داشت که سری زمانی موردنظر WN است. یکی دیگر از ساده‌ترین روش‌های تشخیص WN، رسم نمودار سری زمانی است. در صورت وجود تغییرات در روند سری زمانی، WN نیست و برعکس.

یکی از روش‌های قابل استناد در تشخیص WN، استفاده از نمودار ACF است. در قسمت قبل بیان شد که در صورت قرار گرفتن مقادیر ضریب همبستگی در فاصله‌ی اطمینان، ارتباط و همبستگی میان تأخیرها وجود ندارد و می‌توان گفت که سری زمانی موردنظر تصادفی است. این ویژگی‌ها معرف WN نیز هستند.

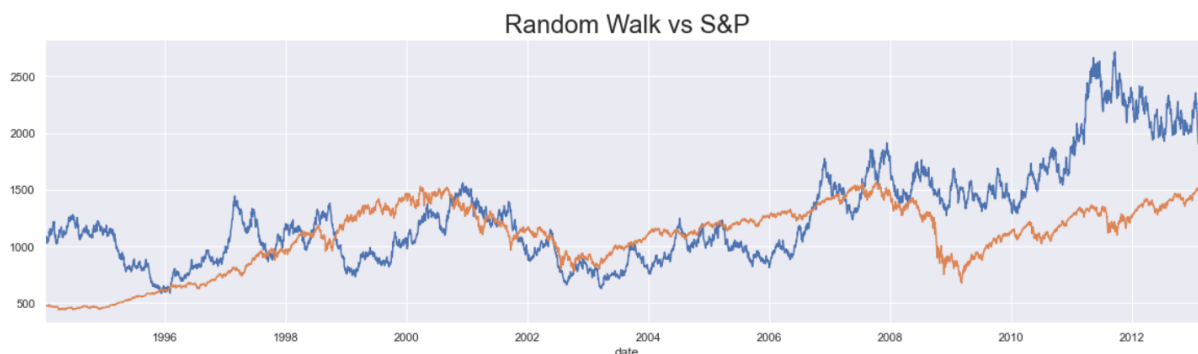


شکل ۱۰-۲: نمودار ACF برای سری زمانی White Noise [۹]

همان طور که در شکل ۲-۸ مشخص شده است، به جز تأخیرهای ۲ و ۳۰ باقی ضرایب همبستگی در فاصله‌ی اطمینان قرار دارند که نشان‌دهنده‌ی عدم وجود همبستگی میان مقادیر این سری زمانی و در نتیجه تصادفی بودن آن است. با توجه به توضیحات ارائه شده، با یک تابع مولد اعداد تصادفی می‌توان یک سری زمانی WN را تولید کرد.

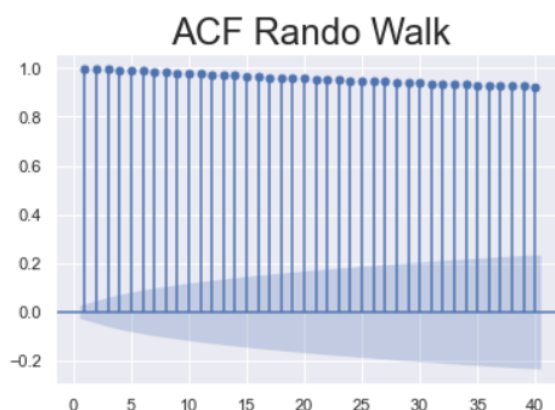
۲-۴-۷ پیاده‌روی تصادفی (Random Walk)

پیاده‌روی تصادفی یا Random Walk نیز نوع دیگری از سری‌های زمانی است. مقادیر آینده‌ی این نوع سری زمانی، مانند WN، قابل پیش‌بینی نیست زیرا همان طور که از نام این سری زمانی پیداست، مقادیر آن به صورت تصادفی تولید می‌شوند. تفاوتی که این نوع سری زمانی با WN دارد، در روش تولید مقادیر آن است. در WN هر مقدار تولیدی در سری زمانی بدون وابستگی به زمان و داده‌های دیگر سری است. ولی در پیاده‌روی تصادفی هر مقدار با توجه به مقدار قبلی تولید می‌شود؛ به این صورت که به مقدار کنونی یک مقدار تصادفی اضافه شده و مقدار بعدی تولید می‌شود. می‌توان با شروع از صفر و یک تابع تولیدکننده‌ی عدد رندوم یک سری زمانی RW تولید کرد.



شکل ۱۱-۲: نمودار یک سری زمانی RW در کنار یک سری زمانی نرمال

برخلاف WN که رسم نمودار یکی از روش‌های تشخیص آن بود، با رسم نمودار RW نمی‌توان آن را تشخیص داد زیرا همان طور که در شکل ۲-۹ نشان داده شده است، نمودار حاصل کاملاً شبیه به یک نمودار سری زمانی نرمال است. در این شکل، روند نارنجی‌رنگ نشان‌دهنده‌ی سری زمانی RW و روند آبی‌رنگ نشان‌دهنده‌ی سری زمانی نرمال است. اگر از نمودار ACF نیز برای تشخیص یک سری زمانی RW استفاده کنیم، نتیجه تفاوتی با یک سری زمانی نرمال ندارد.



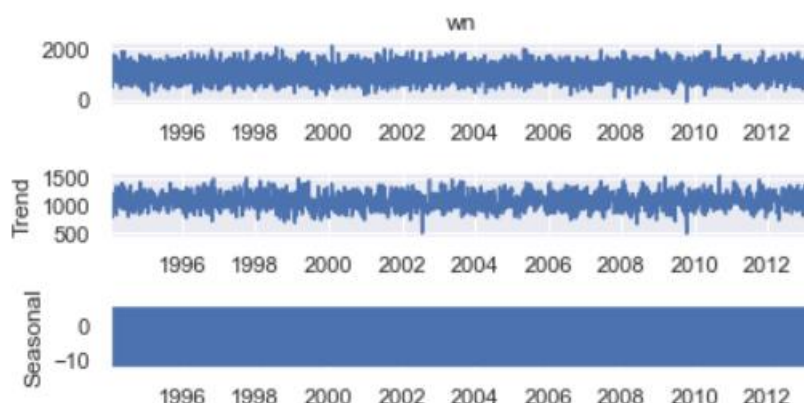
شکل ۱۲-۲: نمودار ACF برای سری زمانی RW

اگر شکل ۲-۵ را با شکل ۲-۱۰ مقایسه کنیم، تفاوت چندانی مشاهده نمی‌شود. یکی از راه‌های تشخیص سری زمانی RW ایستا کردن سری است که در ادامه به توضیح آن پرداخته شده است.

۲-۴-۸ ایستایی (Stationarity)

ایستایی در سری‌های زمانی به معنی عدم وجود تغییرات در کلیت مقادیر داده‌ها در یک بازه‌ی زمانی مشخص است. این مفهوم به معنی عدم تغییر مقدر نیست. به طول مثال white noise طی زمان در حال تولید مقادیر کاملاً تصادفی است ولی نمونه‌ای از سری زمانی ایستا است. سری‌های زمانی ایستا دارای ویژگی‌های مستقل از زمان هستند. میانگین، واریانس در آن‌ها ثابت است و کوواریانس نیز با گذر زمان دچار تغییر نمی‌شود. از طرفی، وجود خاصیت فصلی و یا وجود روند مشخص در سری زمانی نشان‌دهنده‌ی وابستگی مقادیر به زمان است. در نتیجه سری‌های زمانی ایستا روند مشخص و خاصیت فصلی ندارند. عکس این موضوع نیز صادق است؛ به این صورت که سری‌های زمانی که

دارای خاصیت فصلی و یا روند هستند ایستا نیستند. باید به این نکته توجه داشت که سری‌های زمانی ایستا می‌توانند خاصیت تناوبی داشته باشند.



شکل ۱۳-۲: نمودارهای white noise در تحلیل ایستایی

شکل ۲-۷ نمودارهای تحلیل یک سری زمانی white noise را نشان می‌دهد که دارای روند مشخص و خاصیت فصلی نیست. می‌توان گفت که روند آن به صورت خط افقی است.

روش‌های متعددی برای تشخیص ایستایی یک سری زمانی وجود دارد که روش‌های اصلی شامل رسم نمودار سری زمانی، استفاده از نمودار ACF است. همچنین می‌توان به صورت غیرخودکار به محاسبه‌ی میانگین و واریانس سری زمانی در بازه‌های مختلف پرداخت تا در صورت ثابت بودن ایستایی سری زمانی اعلام شود. ولی این روش از نظر بازده غیرقابل استفاده است. اگر در نمودارهای سری زمانی روند و خاصیت فصلی دیده شد، سری زمانی ایستا نیست. همچنین در نمودار ACF اگر مقادیر تأخیرها در ناحیه‌ی اطمینان حضور داشته باشند، سری زمانی ایستا است. در غیر این صورت، اگر مقادیر تأخیرها خارج از ناحیه‌ی اطمینان باشند سری زمانی ایستا نیست.

روش‌های آماری دیگری نیز برای تشخیص ایستایی سری‌های زمانی وجود دارد. روش دیکی-فولر^۱ یکی از این روش‌ها است که جز دسته‌ی آزمون‌های ریشه‌ی واحد^۲ محسوب می‌شود. این دسته آزمون‌ها تعیین می‌کنند که یک سری زمانی تا چه حد به یک روند وابسته است. این دسته آزمون‌ها روش‌ها و گرایش‌های زیادی دارند که روش دیکی-فولر یکی از آن‌ها است. این تست سری زمانی را به عنوان ورودی دریافت می‌کند و خروجی خود را تولید می‌کند. شکل ۲-۸ یک خروجی نمونه از تست دیکی-فولر است. در خروجی این تست، درصدهایی به عنوان معیار نشان داده می‌شوند که در شکل ۲-۸ در سطرها ۵ الی ۷ خروجی قرار دارند. سطر اول خروجی باید با این معیارها مقایسه شود تا درصدی که به آن اندازه احتمال می‌رود داده‌های سری زمانی ایستا باشند شناسایی شود. در شکل ۲-۸ عدد سطر اول منفی یک است که از مقدار معیارهای درصدی داده شده بیشتر است. همچنین در سطر دوم عدد به دست آمده حدوداً چهل و یک صدم است که از عدد معیار یعنی پنج صدم بیشتر است. با توجه به نتایج به دست آمده می‌توان نتیجه گرفت که سری زمانی داده شده ایستا نیست.

^۱ Dickey-Fuller Test

^۲ Unit Root Test

```
sts.adfuller(df.market_value)
✓ 1.1s
(-1.7369847452352478,
 0.41216456967706006,
 18,
 5002,
 {'1%': -3.431658008603046,
  '5%': -2.862117998412982,
  '10%': -2.567077669247375},
 39904.880607487445)
```

شکل ۱۴-۲: نتایج تست دیکی-فولر

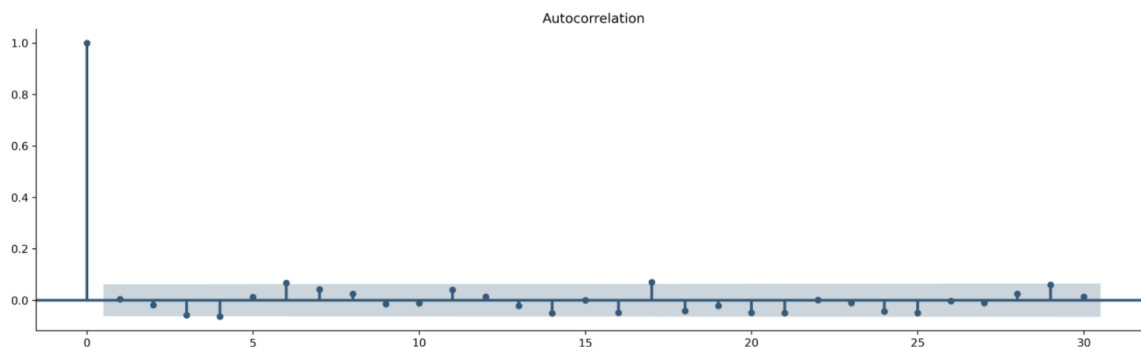
این نکته نیز حائز اهمیت است که روش‌های پیش‌بینی سنتی^۱ سری زمانی قادر به استفاده از سری زمانی غیرایستا نیستند. سری‌های زمانی استفاده شده در این روش‌ها باید بدون روند و خاصیت فصلی باشند زیرا تبدیل کردن آن‌ها به مدل‌های پیش‌بینی سری زمانی آسان است. ولی باید توجه داشت که سری‌های زمانی ایستا که دارای مقادیر کاملاً تصادفی طی زمان هستند را نمی‌توان پیش‌بینی کرد زیرا مقادیر آن‌ها هیچ‌گونه وابستگی‌ای به زمان ندارند. White noise مثالی از این دست سری‌ها است. در صورت شناسایی شدن white noise از هرگونه تحلیل آتی باید جلوگیری شود.

در بخش قبل بیان شد که سری زمانی RW، که یک سری تصادفی است، به راحتی قابل شناسایی نیست و در نمی‌توان آن را پیش‌بینی کرد. همچنین دیدیم که نمودار ACF آن یک سری زمانی غیرایستا را نشان می‌دهد. به طور کلی سری RW، برخلاف سری WN که در ابتدای روند تحلیل شناسایی و متوقف می‌شود، در مرحله‌ی پردازش‌های قبل از پیش‌بینی شناسایی و متوقف می‌شود.

یکی از روش‌های ایستا کردن سری‌های زمانی، تولید تفاوت درجه اول^۲ (FOD) است [۱۱، ۱۲]. برای تولید FOD، مقادیر تأخیرهای سری زمانی را دو به دو از یکدیگر کم می‌کنیم. به عنوان مثال اگر x_1 و x_2 دو تأخیر پشت‌سرهم در یک سری زمانی باشند، $x_2 - x_1$ مقدار اولین تأخیر سری زمانی جدید را تولید می‌کند. اگر FOD یک سری زمانی RW محاسبه شود و نمودار آن به همراه نمودار ACF آن رسم شود، خروجی‌ها نشان‌دهنده‌ی یک سری ایستا مانند WN خواهند بود.

¹ Classic Approaches

² First Order Difference



شکل ۱۵-۲: نمودار ACF سری زمانی RW [۱۲]

همان طور که در شکل ۱۳-۲ نشان داده شده است، به جز موارد معدودی، اکثریت مقادیر تأخیرها در فاصله‌ی اطمینان قرار دارند که نشان‌دهنده‌ی عدم وجود همبستگی میان داده‌های سری زمانی است؛ در نتیجه سری زمانی RW ایستا، غیرقابل پیش‌بینی و تصادفی است.

۲-۵ پیش‌بینی در سری زمانی (Forecasting)

.....Residual.....

۲-۵-۱ مدل‌های پیش‌گویی در سری‌های زمانی

۲-۶ جمع‌بندی

فصل سوم

روش پیشنهادی

۳-۱ مقدمه

۳-۲ جمع بندی

فصل چهارم

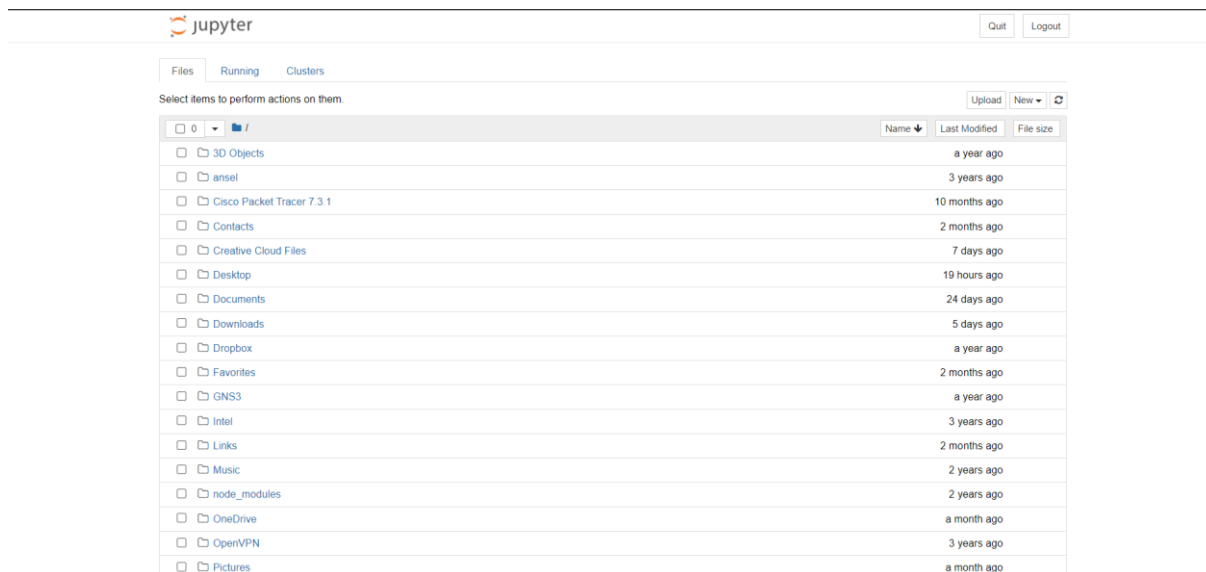
نتایج

۴-۱ مقدمه

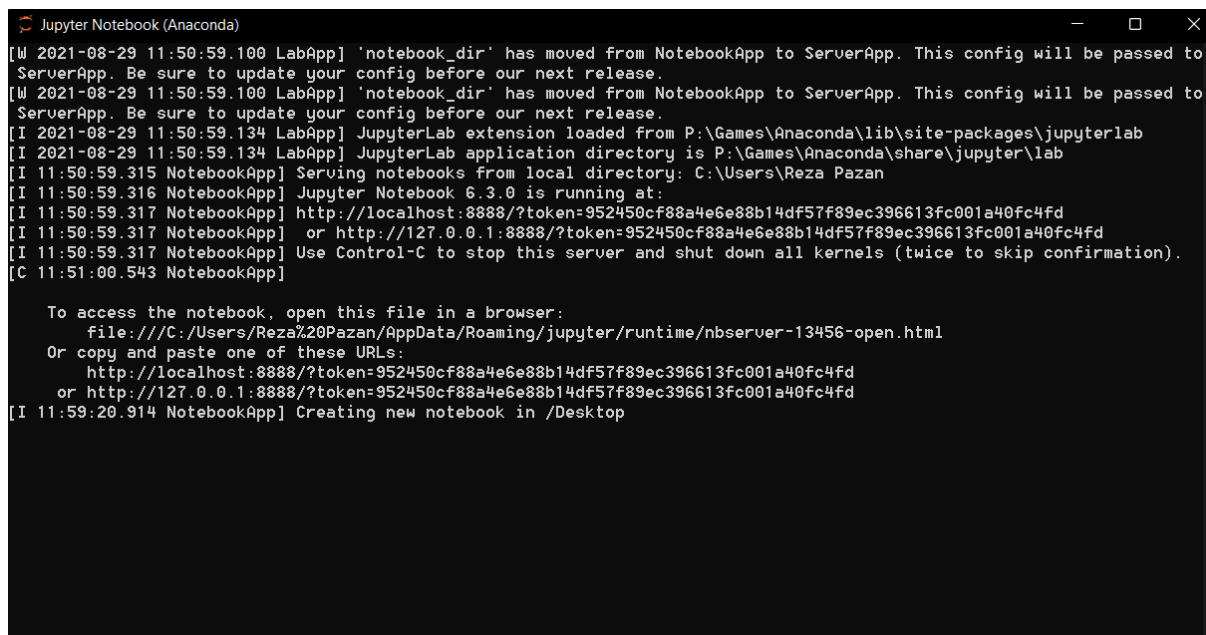
۴-۲ جمع بندی

فصل پنجم

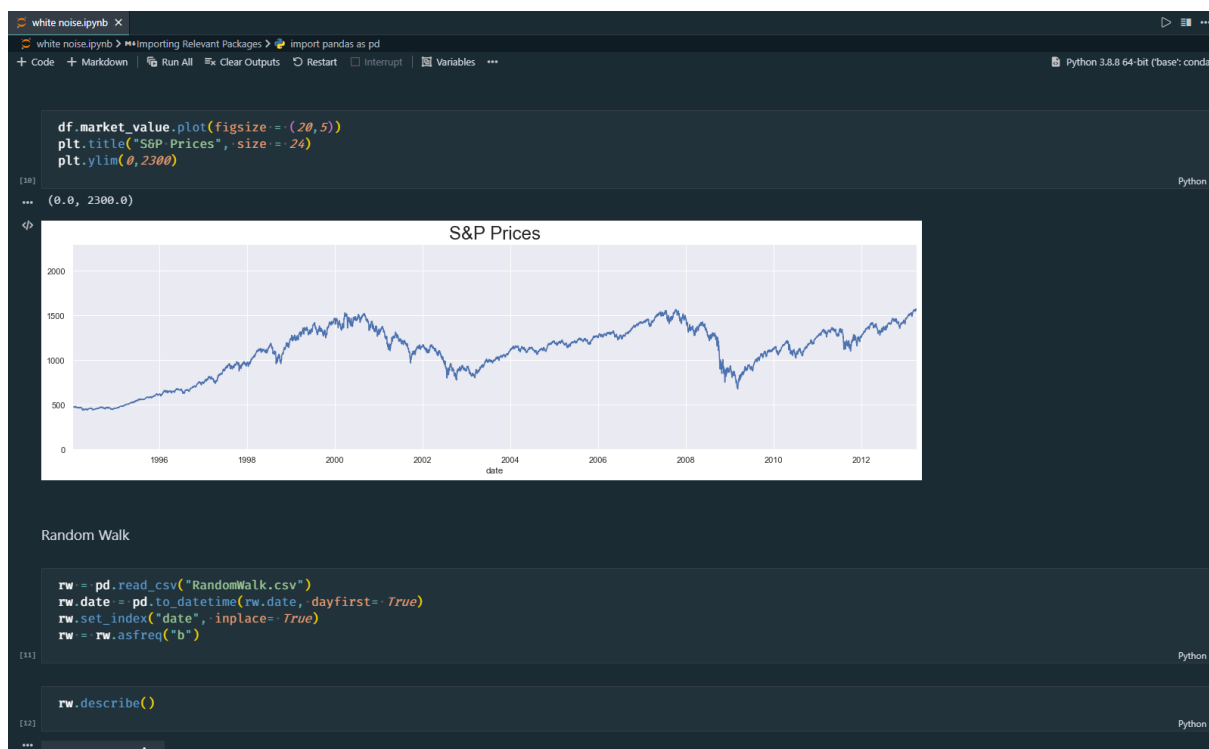
جمع بندی



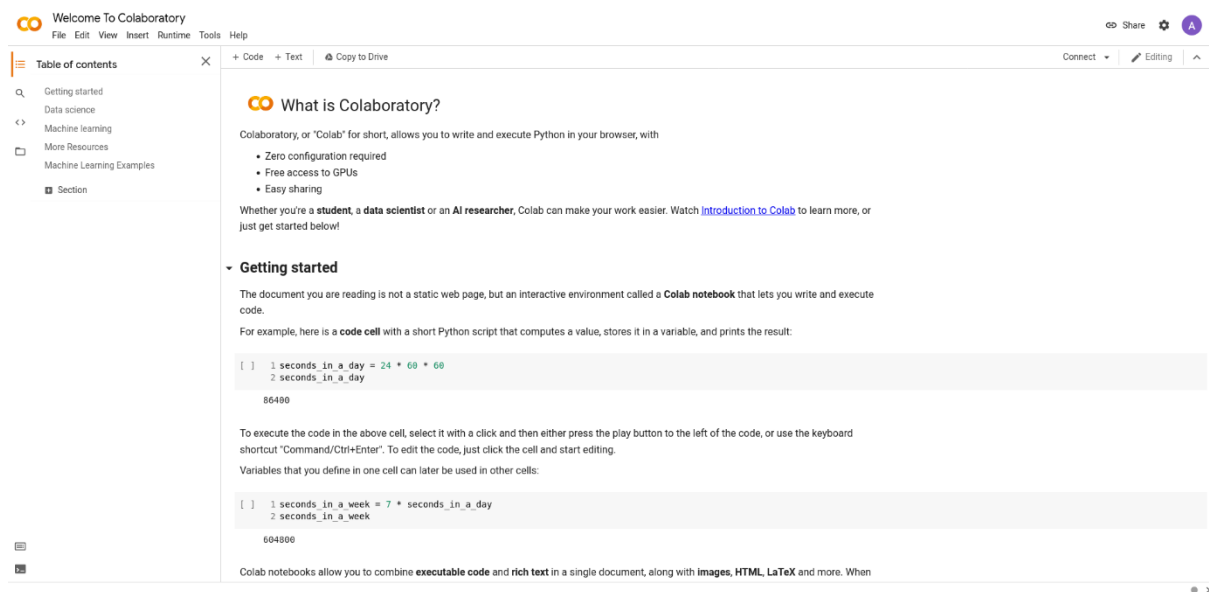
شکل ۱-۶: صفحه‌ی ابتدایی Jupyter Notebook



شکل ۲-۶: سرور اجرا شده برای Jupyter Notebook



شکل ۳-۶: محیط یکپارچه‌سازی شده‌ی Jupyter در Visual Studio Code



شکل ۴-۶: صفحه‌ی ابتدایی Google Colab

- [1] Jung, Sangjoon, Chonggun Kim, and Younky Chung. "A prediction method of network traffic using time series models." In International Conference on Computational Science and Its Applications, pp. 234-243. Springer, Berlin, Heidelberg, 2006.
- [2] Wu, Qingtao, and Zhiqing Shao. "Network anomaly detection using time series analysis." In Joint international conference on autonomic and autonomous systems and international conference on networking and services-(icas-isns' 05), pp. 42-42. IEEE, 2005.
- [3] <https://blog.faradars.org/time-series>, Accessed August 22nd , 2021.
- [4] Joshi, Manish, and Theyazn Hassn Hadi. "A review of network traffic analysis and prediction techniques." arXiv preprint arXiv:1507.05722 (2015).
- [5] Grebennikov, A., Y. Krukov, and D. Chernyagin. "A prediction method of network traffic using time series models." Grebennikov, Y. Krukov, D. Chernyagin.–2011 (2011).
- [6] Kiran, Mariam, Cong Wang, George Papadimitriou, Anirban Mandal, and Ewa Deelman. "Detecting anomalous packets in network transfers: investigations using PCA, autoencoder and isolation forest in TCP." Machine Learning 109, no. 5 (2020).
- [7] Ntlangu, Mbulelo Brenwen, and Alireza Baghai-Wadji. "Modelling network traffic using time series analysis: A review." In Proceedings of the International Conference on Big Data and Internet of Thing, pp. 209-215. 2017.
- [8] Derek Banas, Time Series Analysis, https://www.youtube.com/playlist?list=PLGLfVvz_LVvSVgVCsPWLr961id7kRv1wt, Accessed August 11th, 2021.
- [9] Time Series Analysis in Python 2020 by 365 Careers, <https://www.udemy.com/course/time-series-analysis-in-python>, Accessed August 1st, 2021.
- [10] <https://machinelearningmastery.com/white-noise-time-series-python>, Accessed August 30th, 2021.
- [11] Tony Moses, Udacity, Business Analyst Nanodegree, Time Series Forecasting, <https://git.ir/udacity-time-series-forecasting>.
- [12] Time Series From Scratch, <https://towardsdatascience.com/time-series-from-scratch-white-noise-and-random-walk-5c96270514d3>, Accessed August 27th, 2021.
- [13] Aldhyani, Theyazn HH, Melfi Alrasheedi, Ahmed Abdullah Alqarni, Mohammed Y. Alzahrani, and Alwi M. Bamhdi. "Intelligent hybrid model to enhance time series models for predicting network traffic." IEEE Access 8 (2020): 130431-130451.
- [14] Zhao, Zheng, Weihai Chen, Xingming Wu, Peter CY Chen, and Jingmeng Liu. "LSTM network: a deep learning approach for short-term traffic forecast." IET Intelligent Transport Systems 11, no. 2 (2017): 68-75.