موضوع پروژه:

- با استفاده از اطلاعات موجود در این دیتا ست USA\_Housing.csv قیمت خانه ها رو پیش بینی کنید ؟

خروجی مورد انتظار:

انتظار این است دانشجویان خروجی پروژه را در قالب فایل html تحویل دهند که در آن کدها به همراه خروجی آنها، نمودارها ،

توضیحات و نتیجه گیریها آمده باشد. برای این منظور میتوانید از امکانات Jupyter Notebook استفاده کنید.

انتظار این است براساس مدل DM-CRISP خروجی پروژه شما از بخش های زیر تشکیل شده باشد و به این سوالات پاسخ داده

شده باشد:

نمایش 5 سطر اول وآخر اطلاعات دیتا ست

نمایش اطلاعات مهم دیتا ست

خلاصه ی آماری از داده های عددی

نمایش ستون های دیتا ست

بررسی داده های تکراری در دیتا ست

EDA

نمودار ستونی

Pairplot

متغییر وابسته distplot

متغییر وابسته boxplot

Scatter plot

فهم مسئله

1- انگیزه اصلی چنین پروژهای چیست؟

انگیزه اصلی پروژه‌هایی مانند پیش‌بینی قیمت خانه با استفاده از الگوریتم‌های مانند رگرسیون خطی می‌تواند به چند جنبه برگردد:

پیش‌بینی دقیق: با استفاده از رگرسیون خطی و مدل‌های پیش‌بینی دیگر، می‌توان به صورت دقیق قیمت خانه‌ها را پیش‌بینی کرد. این امر برای خریداران و فروشندگان ملک اهمیت زیادی دارد تا بتوانند تصمیمات بهتری بگیرند.

تحلیل عوامل تأثیرگذار: با استفاده از رگرسیون خطی، می‌توان عواملی که بیشترین تأثیر را بر قیمت خانه دارند را شناسایی کرد. این اطلاعات می‌تواند به تصمیم‌گیری‌های استراتژیک در زمینه خرید و فروش ملک‌ها کمک کند.

بهبود تصمیم‌گیری: پیش‌بینی دقیق قیمت خانه‌ها از طریق رگرسیون خطی می‌تواند به مشتریان کمک کند تا تصمیمات بهتری در مورد سرمایه‌گذاری در بازار مسکن بگیرند و از طریق این پیش‌بینی‌ها ریسک‌های مرتبط با خرید یا فروش ملک را کاهش دهند.

تحلیل و پیش‌بینی بازار: از طریق پروژه‌های مانند پیش‌بینی قیمت خانه، می‌توان بازار مسکن را بهتر درک کرد و تغییرات و روندهای آینده را پیش‌بینی کرد.

به طور کلی، انگیزه اصلی همچین پروژه‌هایی ، بهبود تصمیم‌گیری، افزایش دقت پیش‌بینی، شناسایی عوامل تأثیرگذار و بهبود تحلیل بازار است.

2- خروجی چنین پروژهای برای چه مواردی ممکن است کاربرد داشته باشد؟

به طور کلی، خروجی‌های چنین پروژه‌های می‌تواند در تصمیم‌گیری‌های مرتبط با خرید و فروش ، سرمایه‌گذاری، تحلیل بازار، و تصمیمات سیاستی مورد استفاده قرار گیرد و به بهبود فرآیندها و تصمیمات مرتبط با بازار کمک کند.

اما به طور خاص برای این پروژه

تصمیم‌گیری در خرید و فروش ملک: خروجی این پروژه‌ها می‌تواند به خریداران و فروشندگان کمک کند تا تصمیمات بهتری در مورد خرید و فروش ملک‌ها بگیرند. و قیمت مناسبی را برای ملک‌ها تعیین کنند.

برنامه‌ریزی سرمایه‌گذاری: خروجی‌های این پروژه‌ها می‌تواند به سرمایه‌گذاران و توسعه‌دهندگان املاک کمک کند تا تصمیمات بهتری در مورد سرمایه‌گذاری در بازار مسکن بگیرند و به سمت مناطقی با بازدهی بالاتر حرکت کنند.

تحلیل بازار مسکن: خروجی‌های این پروژه‌ها می‌تواند به محققان و تحلیل‌گران بازار مسکن کمک کند تا روندها و تغییرات بازار را بهتر درک کرده و پیش‌بینی کنند.

تصمیم‌گیری سیاستی: این خروجی‌ها می‌تواند به مقامات شهری و دولتی کمک کند تا تصمیمات سیاستی مناسبی در زمینه مسکن و توسعه شهری بگیرند و بهبود در زمینه سیاست‌گذاری‌های مسکن ایجاد کنند.

3-چه کسانی ممکن است به نتایج این پروژه علاقمند باشند؟ چرا؟

به طور کلی، افراد و نهادهای مختلفی که در جهت کمک به بهبود تصمیم‌گیری‌ها، سرمایه‌گذاری‌ها، تحلیل بازار، و تصمیمات سیاستی مرتبط با بازارهستند علاقمند به نتایج چنین پروژه هایی می باشند

به طور خاص برای این پروژه تحلیل قیمت خانه. افراد و نهادهای زیر ممکن است به نتایج این پروژه علاقمند باشند:

خریداران و فروشندگان ملک: این گروه از افراد به دنبال اطلاعات دقیق و قابل اعتماد در مورد قیمت‌های ملک‌ها هستند تا بتوانند تصمیمات بهتری در مورد خرید و فروش ملک‌ها بگیرند و از طریق این پروژه‌ها قیمت مناسبی را تعیین کنند.

سرمایه‌گذاران و توسعه‌دهندگان املاک: این گروه از افراد به دنبال اطلاعاتی هستند که به آن‌ها کمک کند تا تصمیمات سرمایه‌گذاری مناسبی در زمینه بازار مسکن بگیرند و به سمت مناطق با بازدهی بالاتر حرکت کنند.

مشاوران املاک: این گروه از افراد به دنبال اطلاعاتی هستند که به آن‌ها کمک کند تا به مشتریان خود خدمات بهتری ارائه دهند و تصمیمات منطقی‌تری در زمینه خرید و فروش ملک‌ها بگیرند.

محققان و تحلیل‌گران بازار مسکن: این گروه از افراد به دنبال داده‌ها و اطلاعاتی هستند که به آن‌ها کمک کند تا روندها و تغییرات بازار مسکن را بهتر درک کرده و پیش‌بینی کنند.

مقامات شهری و دولتی: این گروه از افراد به دنبال اطلاعاتی هستند که به آن‌ها کمک کند تا تصمیمات سیاستی مناسبی در زمینه مسکن و توسعه شهری بگیرند و بهبود در زمینه سیاست‌گذاری‌های مسکن ایجاد کنند.

درک داده:

1-داده ها از کجا بدست آمده اند و چگونه جمع آوری شده اند؟

داده‌های مورد استفاده برای همچین پروژه‌های پیش‌بینی قیمت خانه معمولاً از منابع مختلفی جمع آوری می‌شوند. برخی از منابع اصلی جمع آوری داده‌ها عبارتند از:

وب‌سایت‌های املاک: اطلاعات مربوط به فروش و اجاره ملک‌ها از وب‌سایت‌های املاک مختلف جمع آوری می‌شود. این اطلاعات شامل قیمت‌ها، موقعیت جغرافیایی، اندازه ملک، تعداد اتاق‌ها و سایر ویژگی‌های ملک می‌باشد.

سازمان‌های ملکی: برخی از سازمان‌های ملکی و املاکی اطلاعات مربوط به معاملات ملکی را جمع آوری و منتشر

داده‌های عمومی: داده‌های عمومی مانند اطلاعات جمعیتی، اقتصادی، امنیتی، و اطلاعات محیطی

ثبت احوال و اسناد: با دسترسی وسیع به اطلاعات هویتی و جمعیتی

2- هر یک از متغیرها چه چیزی را اندازه گیری می کنند؟

متغیرهای مختلفی استفاده شده است که هر یک از آن‌ها به آیتم خاصی که می تواند بر چیزهای مختلفی را اندازه‌گیری می‌کنند. برخی از متغیرهای معمولاً استفاده شده در این پروژه‌ها عبارتند از:

متغیر قیمت: این متغیر قیمت فعلی یا پیش‌بینی شده ملک را اندازه‌گیری می‌کند. که در این مسئله، target ماست

متغیر‌های مکانی: آدرس، که در این مسئله ، دیتای خاصی را به ما نمی دهد و تاثیری بر قیمت نداشته ، پس آن را نادیده گرفتیم.

متغیر‌های فیزیکی: این متغیرها شامل ویژگی‌های فیزیکی ملک مانند اندازه، تعداد اتاق‌ها، تعداد حمام‌ها، سال ساخت و وضعیت نگهداری ملک می‌باشد. در این مسئله، تعداد اتاق ها و اتاق خواب ها، مدت ساخت خانه را داریم

متغیر‌های اقتصادی: این متغیرها شامل عوامل اقتصادی مانند نرخ بهره، نرخ تورم، درآمد متوسط منطقه و شرایط بازار مسکن است. که در این مسئله، متوسط درامد ساکنین خانه را داریم

متغیر‌های اجتماعی: این متغیرها شامل عوامل اجتماعی مانند جمعیت منطقه، نرخ بیکاری، وضعیت اجتماعی و فرهنگی منطقه می‌باشد. که در اینجا، جمعیت منطقه است

3-آیا ابهامی در تعاریف داده ها وجود دارد؟

داده های مورد نیاز برای پیش بینی قیمت، همگی دارای مقادیر عددی هستند که در غیر این صورت نیاز به تبدیل داده ها داشتیم و دیتای از دست رفته و missing value نداشتیم

البته در تعاریف داده‌ها، بیان مفاهیم دقیق و روشن تر می‌تواند به بهبود درک و استفاده از داده‌ها کمک کند.

مثلا

اگر داده ای نیاز به توضیحات و تفسیر داشته باشد باید آن را تفسیر نمود.

مشخص کردن واحد دقیق اندازه گیری، در صورت داشتن ابهام در مورد واحدهای اندازه‌گیری متغیرها می‌تواند به مشکلات در تحلیل و پیش‌بینی منجر شود.

ابهام در مورد روش‌های جمع آوری داده‌ها و فرآیند تهیه داده‌ها می‌تواند به تردید در کیفیت و قابلیت اطمینان داده‌ها منجر شود.

تعاریف متغیرهای وابسته و مستقل و نحوه رابطه آن‌ها با یکدیگر بهتر است تفسیر شود

4- آیا ممکن است در اندازه گیری متغیرها و یا ثبت داده ها خطایی وجود داشته باشد؟

در پروژه‌های پیش‌بینی قیمت خانه، ممکن است در اندازه‌گیری متغیرها و ثبت داده‌ها خطاهای مختلفی وجود داشته باشد که می‌تواند به تحلیل‌های نادرست و نتایج نادرست منجر شود. برخی از خطاهای ممکن در اندازه‌گیری متغیرها و ثبت داده‌ها عبارتند از:

1. خطاهای انسانی: خطاهای انسانی ممکن است در فرآیند جمع آوری داده‌ها و ثبت آن‌ها رخ دهد. این شامل خطاهای تایپیستی، اشتباهات در ورود داده، یا عدم دقت در ثبت اطلاعات است.
2. خطاهای نمونه گیری: اگر نمونه‌گیری به درستی انجام نشود و نمونه‌ها نماینده جامعه نباشند، ممکن است نتایج تحلیل‌ها نادرست باشد.
3. خطاهای اندازه‌گیری: خطاهای مرتبط با اندازه‌گیری ممکن است در متغیرها وجود داشته باشد. این شامل خطاهای دستگاه اندازه‌گیری، عدم دقت در اندازه‌گیری، یا عوامل خارجی مانند شرایط محیطی است.
4. خطاهای مفقودی: خطاهای مفقودی به معنای عدم ثبت یا گم شدن برخی از داده‌ها است که ممکن است منجر به انحرافات در تحلیل‌ها شود.
5. خطاهای ناشی از تغییرات زمانی: تغییرات در شرایط و زمان ممکن است منجر به خطاهایی در داده‌ها شود که نیاز به تصحیح و تطبیق دارند.

با توجه به این خطاها، اهمیت اعتبارسنجی و پیش‌پردازش داده‌ها بسیار بالاست تا از دقت و قابلیت اطمینان داده‌ها و نتایج حاصل اطمینان حاصل شود.

5- چه متغیرهای دیگری اگر وجود داشتند، میتوانست به حل مسئله کمک کند؟

برای بهبود و دقیق‌تر کردن پروژه پیش‌بینی قیمت خانه، متغیرهای دیگری که اگر وجود داشتند می‌توانستند به حل مسئله کمک کنند، عبارتند از:

متغیر‌های مرتبط با ترافیک و حمل و نقل عمومی: اضافه کردن متغیرهایی مانند فاصله به ایستگاه‌های حمل و نقل عمومی، زمان سفر به مراکز شهری، و امکانات حمل و نقل ممکن است به بهبود پیش‌بینی قیمت خانه کمک کند.

متغیر‌های مرتبط با امکانات محله: اضافه کردن ویژگی‌های محله مانند امکانات ورزشی، رستوران‌ها، فضای سبز، و امکانات فرهنگی می‌تواند به تحلیل بهتر و دقیق‌تر اطلاعات منطقه کمک کند.

متغیر‌های مرتبط با امنیت و جرم: اضافه کردن متغیرهایی مانند نرخ جرم و تعداد پلیسی که در منطقه فعالیت دارند، می‌تواند به تحلیل امنیت منطقه و تأثیر آن بر قیمت خانه کمک کند.

متغیر‌های مرتبط با تغییرات اقلیمی: اضافه کردن متغیرهایی مانند دما، بارش، و شرایط هوا و هوایی می‌تواند به تحلیل تأثیر تغییرات اقلیمی بر قیمت خانه کمک کند.

متغیر‌های مرتبط با تحولات اقتصادی: اضافه کردن متغیرهایی مانند نرخ بهره، نرخ تورم، و شرایط اقتصادی می‌تواند به تحلیل تأثیرات اقتصادی بر بازار مسکن کمک کند.

با اضافه کردن این متغیرها و تحلیل دقیق آن‌ها، می‌توان به تحلیل بهتر و دقیق‌تری از قیمت خانه‌ها دست یافت.

6- متغیرهای موجود از کدام نوعند )رسته ای – عددی( ؟

متغیرهای موجود در پروژه پیش‌بینی قیمت خانه می‌توانند به دو دسته اصلی تقسیم شوند: متغیرهای رسته‌ای و متغیرهای عددی.

1. متغیرهای رسته‌ای: این نوع متغیرها معمولاً دارای مقادیر گسسته و محدودی هستند و به عنوان عوامل تصمیم‌گیری مورد استفاده قرار می‌گیرند. مثال‌هایی از متغیرهای رسته‌ای در این پروژه می‌تواند شامل ویژگی‌هایی مانند نوع ملک (آپارتمان، ویلا، زمین و غیره) یا وضعیت نگهداری (نوساز، دست دوم و غیره) باشد.
2. متغیرهای عددی: این نوع متغیرها دارای مقادیر عددی پیوسته هستند و معمولاً به عنوان ورودی‌های محاسباتی در مدل‌های پیش‌بینی استفاده می‌شوند. مثال‌هایی از متغیرهای عددی در این پروژه می‌تواند شامل قیمت فعلی یا پیش‌بینی شده ملک، تعداد اتاق‌ها، سن ساخت، و غیره باشد.

با توجه به این تقسیم‌بندی، متغیرهای رسته‌ای معمولاً نشان دهنده ویژگی‌های کیفی ملک هستند، در حالی که متغیرهای عددی اطلاعات کمی و محاسباتی را نشان می‌دهند که می‌توانند برای پیش‌بینی قیمت خانه مورد استفاده قرار بگیرند.

که در اینجا متغیرهایی که در تحلیل استفاده می شود، همگی از نوع عددی بوده، بجز آدرس که تاثیری در قیمت نداشت و نادیده گرفتیم

7- خلاصه آماری متغیرهای موجود چیست؟

میانگین: میانگین یک متغیر عددی نشان دهنده مقدار متوسط آن متغیر است و با محاسبه مجموع تمام مقادیر و تقسیم بر تعداد مقادیر محاسبه می‌شود.

واریانس و انحراف معیار: واریانس و انحراف معیار نشان دهنده پراکندگی داده‌ها از میانگین می‌باشند. واریانس معمولاً محاسبه شده بر اساس میانگین مربعات اختلاف هر مقدار از میانگین است.

حداقل و حداکثر: این آمارها نشان دهنده کمترین و بیشترین مقدار موجود در داده‌ها هستند.

کوارتیل‌ها: کوارتیل‌ها نشان دهنده مقادیری هستند که داده‌ها را به چهار بخش مساوی تقسیم می‌کنند. مثلاً کوارتیل اول نشان دهنده مقداری است که 25 درصد داده‌ها کمتر از آن هستند.

هیستوگرام: هیستوگرام نموداری است که توزیع فراوانی مقادیر مختلف یک متغیر را نشان می‌دهد و می‌تواند به درک بهتر توزیع داده‌ها کمک کند.

آماده سازی داده

1-آیا نیاز به درآمیختن داده ها است؟ اقدامات و نتایج گزارش شود.

از میان متغیر های موجود ، متغیرهای مستقل که در پیش بینی قیمت تاثیر دارند عبارتند از، تعداد اتاق ها، تعداد اتاق خواب، مدت سال و سن ساختمان، موسط درامد، جمعیت ، که همگی تاثیر گذار هستند پس باید باهم ترکیب و با توجه به مقادیر بسیار متفاوتی که دارند باید استاندارد سازی شود که با fit transform این عمل انجام شد.

داده های موجود، متغیر قیمت متغیر وابسته بوده و بقیه متغیرها ، از نوع مستقل هستند، و قیمت در اثر

2-آیا نیاز به پاکسازی داده است؟ اقدامات و نتایج گزارش شود.

با توجه به بررسی های انجام شده و تحلیل های آماری اجرا شده، خروجی plot ها، داده ها نرمال بوده و توزیع نرمالی دارند و چولگی به سمت چپ و راست نداریم

3-آیا نیاز به تبدیل داده است؟ اقدامات و نتایج گزارش شود.

با بررسی داده ها، missing value نداشته و همه داده ها بجز آدرس، از نوع float و عددی بوده و نیاز به تبدیل نداریم. بجز آدرس، که به دلیل آنکه دیتای خاصی در مورد قیمت به ما نمی دهد که نادیده گرفتیم

4- آیا نیاز به کاهش داده است؟ اقدامات و نتایج گزارش شود.

از میان ستون های داده شده ، آدرس، از آنجا که داده خاصی برای پیش بینی به ما نمی دهد ، چشم پوشی می شود. متغیر وابسته قیمت خانه است وبقیه متغیرها مستقل هستند که باید آن ها را در پیش بینی قیمت در نظر گرفت و برای تجمیع آنها نیاز به استاندارد سازی آن هاست

مدلسازی:

روی داده های آموزش ،الگوریتم رگرسیون ساخته شود .اقدامات و نتایج گزارش شود.

دیتاها را به دو قسمت آموزش و آزمایش تقسیم کردیم. 60 درصد را آموزش و 40 درصد را برای آزمایش در نظر گرفتیم و سپس آنها را استاندارد کردیم چون داده ها با اندازه های مختلف بودند، تعداد اتاق ها مثلا 6 و درامد 50000 که نیاز به استاندارد سازی دارند که با fit transform این عمل انجام شد.

سپس مدل رگرسیون خطی را روی آن پیاده کردیم. و با تابع fit آن را آموزش دادیم و از آنحایی که رابطه y=b0+b1X را داریم مقادیر b0 و b1 را محاسبه کردیم

ارزیابی:

1-مدلهای ارائه شده، روی دادههای آزمایش با استفاده از شاخصهای ارزیابی رگرسیون خطی در یادگیری ماشین ارزیابی شوند .اقدامات و نتایج گزارش شود.

جهت بررسی اینکه پیش بینی مدل به خوبی انجام شده یا خیر، از دستور predict و متغیرهای x را به آن دادیم و y\_hat ها را برای ما پیش بینی میکند و باید آن ها را با yهای تست بررسی و مقایسه می کنیم، که تقریبا نموداری با زاویه 45 داریم و نشان میدهد مدل خوبی است.

سپس خطا را محاسبه می کنیم

2-چه پیشنهاداتی دارید تا نتایج در محیط واقعی، آزمایش گردد؟

برای بررسی و آزمایش در محیط واقعی:

جمع آوری داده ها و هر گونه اطلاعاتی که با توجه به منطقه هدف، در قیمت خانه تاثیر گذار است. که می توان آیتم های بیشتری مثل منطقه، دسترسی به مکان ها وامکانات شهری و حمل و نقل و .. باشد

آزمایش در یک محیط واقعی و با آیتم ها و داده های بدست آمده از محیط و در نظر گرفتن شرایط واقعی، مثلا از مدل خود برای پیش بینی قیمت چند خانه جدید استفاده کنیم و نتایج را با واقعیت مقایسه کنیم

آزمون را در محیط های مختلف و مناطق جغرافیایی متفاوت آزمایش کنیم

روش خود را با سایر روش ها و مدل های موجود دیگری که می شود استفاده کرد مقایسه کنیم

استقرار:

حال اگر بخواهید چنین الگوریتمی را در مقیاس صنعتی توسعه دهید، به این فکر کنید با چه چالش هایی مواجه خواهید شد وبرای آن چه راهکارهایی دارید. موارد زیر را گزارش کنید:

1-چالشهای توسعه الگوریتم را بررسی کنید.

مقیاس پذیری: اجرای الگوریتم بر روی مجموعه داده‌های بزرگ و در مقیاس صنعتی ممکن است به چالش مواجه شود.

مدیریت داده: مدیریت و پردازش داده‌های حجیم و پیچیده می‌تواند یک چالش مهم باشد.

پایداری و عملکرد: اطمینان از پایداری و عملکرد الگوریتم در مقیاس صنعتی از جمله چالش‌های مهم است.

2- چه راهکارهایی برای حل آن ها دارید؟

* استفاده از پردازش موازی و پارالل برای افزایش سرعت اجرا و مقیاس پذیری.
* بهینه‌سازی الگوریتم‌ها برای اجرا بهینه‌تر و سریع‌تر در مقیاس بزرگ.
* استفاده از فناوری‌های جدید و مدرن مانند سیستم‌های توزیع شده و ابری برای مدیریت داده‌ها و اجرای الگوریتم

3-چه ملزوماتی برای ارائه آن راهکارها نیاز دارید؟

* زیرساخت قوی: نیاز به زیرساخت قوی و قابل اعتماد برای پردازش داده‌های حجیم.
* توانایی محاسباتی: دسترسی به توان محاسباتی کافی برای اجرای الگوریتم در مقیاس صنعتی.
* تیم متخصص: تیمی از متخصصان در زمینه علوم داده و مهندسی نرم‌افزار برای توسعه، پیاده‌سازی و پشتیبانی الگوریتم.

نتیجه گیری:

1- انجام این پروژه چه یادگیری برای شما داشت؟

1. تحلیل داده: تجربه تحلیل داده‌های مختلف و استفاده از آن‌ها برای استخراج الگوها و اطلاعات مفید برای پیش‌بینی قیمت خانه و پروژه های مشابه
2. پیش‌بینی و مدل‌سازی: آشنایی با روش‌ها و الگوریتم‌های مختلف مدل‌سازی و پیش‌بینی، از جمله استفاده از مدل‌های یادگیری ماشین برای پیش‌بینی قیمت خانه.
3. مدیریت پروژه: تجربه مدیریت زمان و منابع برای انجام پروژه بهینه و موفق.
4. تفکر تحلیلی: توانایی تحلیل مسائل پیچیده و پیدا کردن راه‌حل‌های بهینه برای آن‌ها.
5. مهارت‌های فنی: بهبود مهارت‌های فنی در زمینه علوم داده، یادگیری ماشین، و تحلیل آماری.
6. مدیریت چالش‌ها: تجربه مواجهه با چالش‌های مختلف در طول پروژه و یادگیری راهکارهای موثر برای حل آن‌ها.

به طور کلی، انجام این پروژه به من کمک کرده است تا مهارت‌ها و دانش‌های مختلفی را تقویت کنم و بهبود یابم تا در پروژه‌های آینده بهتر عمل کنم.

2-با چه چالشهایی مواجه شدید؟ چگونه آنها را حل کردید؟

در طول انجام این پروژه، با چالش‌های مختلفی روبرو شدم که برخی از آن‌ها عبارت بودند از:

مواجهه با داده‌های حجیم و پیچیده که نیازمند پردازش و تحلیل دقیق بودند.

* + راه‌حل: برای حل این چالش، از ابزارهای مدیریت داده مانند Pandas و NumPy در Python استفاده کردم. همچنین از تکنیک‌های بهینه‌سازی مانند انتخاب ویژگی‌های مهم و کاهش ابعاد استفاده کردم.

انتخاب مدل مناسب برای پیش‌بینی قیمت خانه از میان مجموعه گسترده‌ای از الگوریتم‌ها یک چالش بود.

* + راه‌حل: برای حل این چالش، از ارزیابی مدل‌های مختلف و تست عملکرد آن‌ها بر روی داده‌های آزمون استفاده کردم. همچنین از تکنیک‌های ارزیابی مدل مانند ارزیابی متقاطع (Cross-Validation) استفاده کردم.

تفسیر و تبیین دقیق عوامل موثر در پیش‌بینی قیمت خانه توسط مدل‌های پیچیده یادگیری ماشین یک چالش بود.

* + راه‌حل: برای حل این چالش، از روش‌های تفسیری مانند شاخص‌های ویژگی مهم (Feature Importance) و تحلیل شبکه‌های عصبی استفاده کردم تا عوامل موثر در تصمیم‌گیری مدل‌ها را بررسی کنم.