

چکیده

مسئله اصلی این پژوهش، بررسی روش های مدرن با افزایش قطعیت در پیش‌نگری خرید و فروش انواع بازار و تبدیل آن ها به تصمیمات عملیاتی است. در بسیاری از بازار ها و مارکت ها (که در اینجا به طور ویژه بیشتر به فروشگاه های خرده فروشی) نوسانات فصلی، تخفیف ها و عوامل بیرونی موجب خطاهای پیش‌بینی و در نتیجه کمبود یا مازاد موجوی می‌شوند که هزینه های قابل توجهی را به همراه دارند.

این پروژه با تمرکز بر روی انجام تحلیل در مجموعه داده های دیجیتال، رویکردی مبتنی بر یادگیری ماشین را پیشنهاد می‌دهد که شامل مهندسی و استخراج ویژگی جامع، نمایش اثرات تبلیغات و ویژگی‌های زمانی فروشگاه، آموزش مدل تقویت شده‌ی درختی (XGBoost) همراه با اعتبارسنجی زمانی است. ابزار ها و فناوری های اصلی شامل پکیج و کتابخانه های Python مانند `XGBRegressor`, `sklearn`, `pandas` و روش های ارزیابی مبتنی بر پنجره های زمانی است. خروجی سیستم؛ پیش‌بینی های روزانه، اهمیت ویژگی ها برای تبیین نتایج و معیار های عملکرد کسب و کار (MAE, WMAPE) و ارزیابی کاهش ریسک کمبود کالا) است. نتایج اولیه نشان می‌دهند که رویکرد پیشنهادی به طور معناداری خطای پیش‌بینی را نسبت به پایه های ساده کاهش می‌دهد و قابلیت تبدیل به تصمیمات تامین و برنامه‌ریزی بازاریابی را دارد.

تعریف مسئله

○ شرح دقیق مسئله تجاری

فروشگاه ها و خرده فروشی ها با نوسانات فصلی، تاثیرات تبلیغات، تغییرات قیمتی و متغیر های خارجی (مثل آب‌وهوا یا رقبا) روبه‌رو هستند؛ این عدم قطعیت باعث کمبود موجودی یا فزونی موجودی می‌شود که هزینه های مستقیم و فرصت‌سوزی به وجود می‌آورد. نیاز تجاری به طور واضح دربرگیرنده داشتن پیش‌بینی های روزانه/هفتگی قابل که بتوانند تصمیمات تامین، تبلیغات و برنامه‌ریزی عملیاتی را بهینه کنند و در نتیجه هزینه نگهداری کالا و زیان ناشی از کالای تمام شده را کاهش دهند، است.

○ چالش های موجود در روش های سنتی

روش های کلاسیک شامل مدل های حوزه سری زمانی (ARIMA, Holt-Winters) و رگرسیون های خطی اند. این روش ها در شرایط پایدار و با الگو های فصلی روشن خوب عمل می‌کنند ولی در مواجهه با چندین متغیر خارجی، تعاملات پیچیده، داده های پرت یا ساختار های سلسله مراتبی (فروشگاه ها × کالا ها) محدودیت دارند. همچنین روش های سنتی دیگر کمتر قادر به استفاده از منابع داده‌ی حجیم و متنوع (پروموشن ها، متادیتای فروشگاه، رقابت) بوده و به ندرت عدم قطعیت یا بازه پیش بینی را به خوبی مدل می‌کنند.

○ هدف پروژه/ چه چیزی بهبود می‌یابد

مقصود پروژه ایجاد یک خط لوله عملیاتی پیش بینی است که با استخراج ویژگی های زمانی و عملیاتی (لاگ ها، میانگین های متحرک، مشخصات فروشگاه) و استفاده از مدل های یادگیری ماشین مشخص (XGBoost) به عنوان مدل اصلی و Random Forest به عنوان بنچمارک)، دقت پیش بینی را نسبت به پایه های ساده به نحو بارز افزایش دهد. خروجی

ها شامل پیش بینی های انجام شده، بازه های اطمینان و ارزیابی و مکانیزم ها قابل تفسیر هستند تا تصمیم گیران تجاری بتوانند پیش بینی ها را به اقدامات عملی (مثل تعیین نقطه سفارش، تخصیص موجودی مناسب و برنامه ریزی تبلیغات) تبدیل کنند که نتیجتاً به کاهش هزینه های موجودی و بهبود درآمد عمل گرای شرکت منجر خواهد شد.

مرور مطالعات و کار های مرتبط

○ معرفی راه حل های موجود و مقایسه با پروژه ما

- در حوزه پیش بینی خرده فروشی، روش ها و سیستم های متنوعی طی چند دهه توسعه یافته اند که هر یک مزایا و محدودیت های خاص خود را دارند. از روش های سنتی می توان به مدل های سری زمانی مانند ARIMA و هموارسازی نمایی (Holt-Winters) اشاره کرد که در بسیاری از مطالعات اولیه به عنوان ابزار های پایه برای تحلیل روند و فصلی بودن فروش به کار رفته اند و پایه ای قابل اطمینان برای برنامه ریزی میان مدت فراهم می کنند ([Wikipedia](#)). همچنین تحلیل ترکیبی سری زمانی و رگرسیون برای پیش بینی تقاضا در صنعت دارو سازی گزارش شده است که در آن اثر تاریخ، پروموشن و قیمت به صورت همزمان مدل شده اند ([DergiPark](#)).
- در سطح پروژه های کاربردی، مطالعات مقایسه ای از مدل های مختلف یادگیری ماشین مانند رگرسیون خطی، الگوهای سلسله مراتبی و تصمیم درختی برای پیش بینی فروش گزارش شده اند. به عنوان نمونه، یک تحقیق کاربردی مدل های درختی را در مقایسه با روش های آماری سنتی بر روی داده های فروش بررسی کرده و گزارش می دهد که این رویکرد ها در مواجهه با عدم پیوستگی داده ها و متغیر های دسته ای عملکرد بهتری ارائه می دهند ([Arxiv](#)). مطالعه ای دیگر نیز نشان می دهد که ترکیب روش های سری زمانی و رگرسیون در ساختار های هیبریدی می تواند دقت پیش بینی را در محیط های با تغییرات بالا بهبود دهد ([OUCL](#)).
- در کنار ادبیات علمی، سیستم های تجاری مدیریت زنجیره تامین و برنامه ریزی فروش مانند Oracle Demantra، SAP APO و SAS Forecasting Studio نیز در صنعت برای پیش بینی فروش و مدیریت تقاضا استفاده می شوند که معمولاً از مدل های سری زمانی و رگرسیون به عنوان بخشی از فرآیند پیش بینی بهره می برند، هرچند تضاد های پیچیده تر بین عوامل مختلف کسب و کار را به طور کامل مدل نمی کنند ([Oracle Docs](#)).
- در مقایسه با این راه حل های موجود، پروژه پیشنهادی تمرکز خود را از مدل های ایستا و قواعد ثابت به سمت مدل سازی داده محور و مقیاس پذیر منتقل می کند. در حالی که روش های سنتی عمدتاً روابط ساده و از پیش فرض شده را مدل می کنند، رویکرد جدید امکان یادگیری الگوهای پیچیده، تعاملات چندگانه میان متغیر ها و سازگاری با تغییرات رفتاری بازار را فراهم می سازد. علاوه بر این، پروژه به جای ارائه صرف پیش بینی عددی، بر تولید خروجی های قابل تفسیر و قابل استفاده در تصمیمات عملیاتی تمرکز دارد؛ موضوعی که در بسیاری از راه حل های قدیمی به صورت محدود پشتیبانی می شود.

روش پیشنهادی

○ معماری کلی سیستم

- **لایه داده:** منابع داده شامل داده های تاریخی فروش، تقویم/تعطیلات، اطلاعات فروشگاه، تاریخچه پروموشن و داده های خارجی. داده ها به صورت منظم در یک مخزن (فایل های CSV / Parquet) نگهداری می شوند.
- **لایه پردازش و ویژگی سازی:** مراحل پاک سازی، همگام سازی زمانی و تولید ویژگی های زمانی و ویژگی های دسته ای/پرومو انجام می شود.
- **لایه مدل سازی:** مدل مقایسه ای (Random Forest) و مدل اصلی (XGBoost) اجرا، اعتبارسنجی زمانی و ارزیابی می شوند.
- **لایه XAI و Monitoring:** مدل های نهایی به صورت توضیح پذیر (XAI) یا خروجی های دسته ای برای داشبورد صادر می شوند؛ شامل گزارش خطا ها و نمایش feature importance.

○ مراحل کار

- I. **جمع آوری داده :** خواندن فایل های حاوی اطلاعات فروش، فروشگاه، تقویم و پروموشن.
- II. **پاک سازی:** اصلاح داده های ناقص، حذف و مدیریت روز های بسته، همگام سازی شاخص های زمانی.
- III. **ویژگی سازی:** ساخت متغیرهای تاریخ، لاگ ها، میانگین های متحرک (30/7 روز)، پرومو-فلگ ها و ویژگی های مربوط به رقبا/فاصله.
- IV. **مدل سازی پایه و مقایسه:** اجرای مدل های پایه (naive seasonal, ARIMA) و مدل مقایسه ای Random Forest
- V. **آموزش مدل اصلی:** آموزش XGBoost با early-stopping و تنظیمات پایه روی folds زمانی.
- VI. **تفسیر و ارزیابی:** محاسبه MAE, RMSE, WMAPE و استخراج feature importance و SHAP برای توضیح پذیری.
- VII. **مانیتورینگ و نمایش:** بسته بندی مدل، ساخت XAI برای نمایش معنادار و ایجاد داشبورد گزارش دهی.

○ ورودی ها، پردازش ها و خروجی ها

• ورودی ها:

- تاریخچه فروش روزانه (Store, Date, Sales, Customers)
- اطلاعات فروشگاه (StoreType, Assortment, CompetitionDistance)
- تقویم و تعطیلات (StateHoliday, SchoolHoliday)
- تاریخچه پروموشن/قیمت

- پردازش ها:

- پاک سازی و پر کردن شکاف ها
- تولید لاگ ها و کدنویسی دسته ای
- مقیاس بندی و آماده سازی برای مدل
- اعتبارسنجی زمانی و محاسبه متریک ها

- خروجی ها:

- پیش بینی روزانه فروش (point forecasts)
- معیار های عملکرد (MAE, RMSE, WMAPE) و گزارش های خلاصه
- feature importance / SHAP summaries برای هر فروشگاه و گروه

- منطق اصلی سیستم و الگوریتم ها

- رویکرد اصلی: مدل سازی نقطه ای با **XGBoost** روی مجموعه ای از ویژگی های مهندسی شده از سری زمانی و متغیر های عملیاتی.

- چرا منطق **XGBoost**: توانایی یادگیری روابط غیرخطی و تعاملات بین متغیر های گوناگون (تقویمی، پروموشن، ویژگی های فروشگاه) را دارد، در عین حال آموزش و استقرار آن نسبتا سریع و مهندسی شدنی است.

- الگوریتم (خلاصه):

- i. برای هر نمونه (store, date) برداری از ویژگی ها تولید می شود.
- ii. داده ها طبق تاریخ به folds زمانی تقسیم می شوند.
- iii. مدل **XGBoost** با loss مناسب روی هر fold آموزش و با early stopping اعتبارسنجی می شود.
- iv. خروجی ها تجمیع و با میانگین یا stack ساده ترکیب می شوند.
- v. برای توضیح پذیری از SHAP یا feature importance استفاده می شود.

- بنچمارک: یک مدل درختی ساده (Random Forest) و یک baseline سری زمانی اجرا می شوند تا ارزیابی سود واقعی **XGBoost** قابل استناد باشد.

ابزارها و فناوری های مورد استفاده

○ زبان برنامه نویسی

- زبان Python به عنوان زبان اصلی به دلیل پشتیبانی گسترده از تحلیل داده، یادگیری ماشین و سادگی پیاده سازی.

○ فریمورک ها و نرم افزار ها

- محیط Jupyter Notebook و Google Collab برای توسعه، آزمایش و مستندسازی مدل ها.

○ پایگاه داده

- [Rossmann Store Sales](#)

○ کتابخانه ها

- **Scikit-learn** برای پیش پردازش داده ها، مدل های پایه و ارزیابی.
- **XGBoost** برای آموزش مدل پیش بینی فروش.
- **SHAP (SHapley Additive exPlanations)** برای تحلیل اهمیت ویژگی ها و تفسیر خروجی مدل ها در سطح کلی و فروشگاه محور.
- **Feature Importance** داخلی مدل های درختی برای ارائه توضیحات سریع و قابل فهم به کاربران تجاری.

پیاده سازی انجام شده تا میان ترم

○ بخش هایی که تاکنون پیاده سازی شده اند

- طبق برنامه ریزی بخش "روش پیشنهادی"، مراحل اصلی به صورت گام به گام پیاده سازی شده است. در ابتدا داده های آموزش، آزمون و اطلاعات فروشگاه ها بارگذاری و یکپارچه سازی شدند. سپس تحلیل اکتشافی داده ها (EDA) برای درک رفتار فروش، اثر تعطیلی ها، پروموشن ها و الگو های زمانی پیاده سازی گردید. در ادامه پیش پردازش داده ها شامل مدیریت مقادیر گم شده، تبدیل متغیر های دسته ای، استخراج ویژگی های زمانی و ساخت ویژگی های مبتنی بر تاریخ انجام شد. در نهایت، داده ها برای آموزش مدل ها آماده سازی و تقسیم بندی شدند؛ پس از آموزش ارزیابی با متریک های مناسب و مقایسه مدل اصلی با مدل بنچمارک انجام شد. همچنین بخشی از قسمت های اولیه XAI را طراحی کرده ایم.

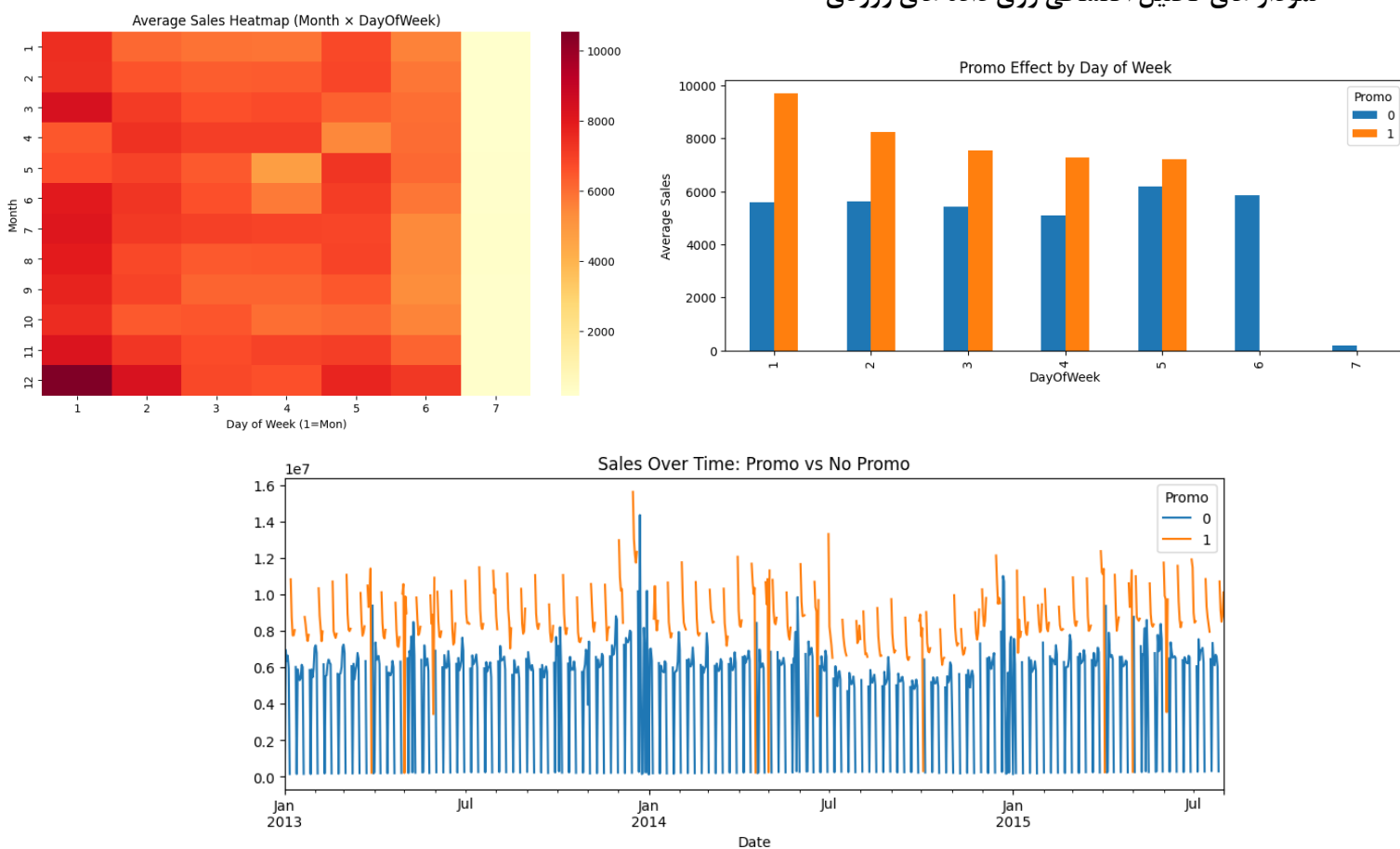
○ توضیح عملکرد بخش های پیاده سازی شده

- عملکرد سیستم به گونه ای طراحی شده که ابتدا الگو های تاریخی فروش و عوامل مؤثر بر آن شناسایی شوند. ویژگی های مهندسی شده به مدل اجازه می دهند اثر زمان، تعطیلی فروشگاه، پروموشن و روند های فصلی را در نظر بگیرد. سپس با استفاده از مدل یادگیری ماشین مبتنی بر (XGBoost)، فرآیند آموزش مدل انجام شده و از مکانیزم توقف زودهنگام (Early Stopping) برای جلوگیری از بیش برآزش استفاده شده است. خروجی مدل شامل پیش بینی فروش و ارزیابی عملکرد آن بر روی داده های اعتبارسنجی است.

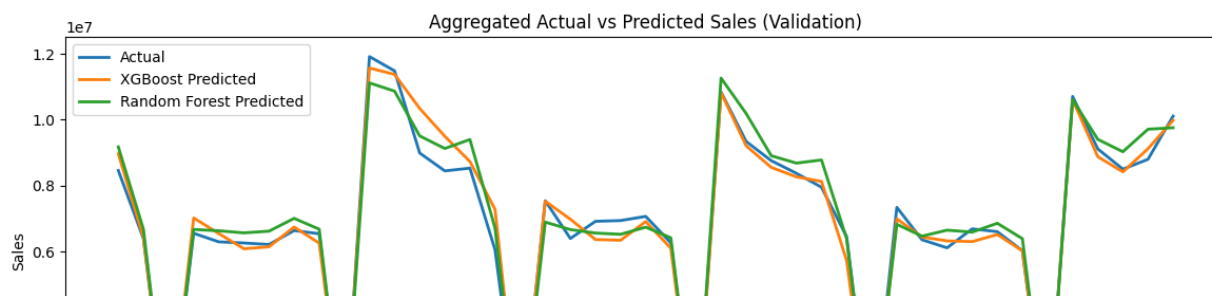
روال کامل انجام کار، ساختار کد ها، خروجی ها و ترتیب مراحل پیاده سازی به صورت شفاف در مخزن [GitHub](#) پروژه مستند شده است تا امکان بررسی جزئی تر و توسعه پروژه برای سایر افراد فراهم باشد.

نتایج اولیه و تحلیل

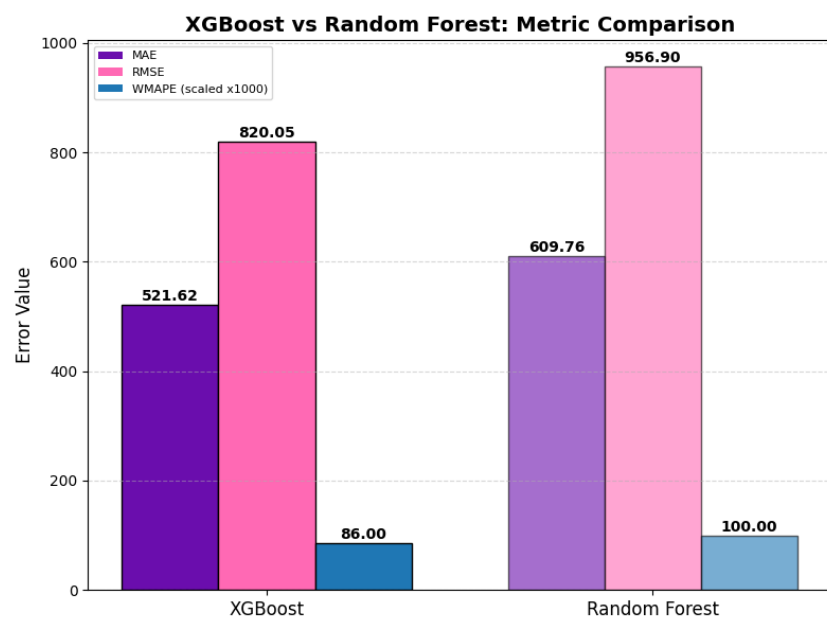
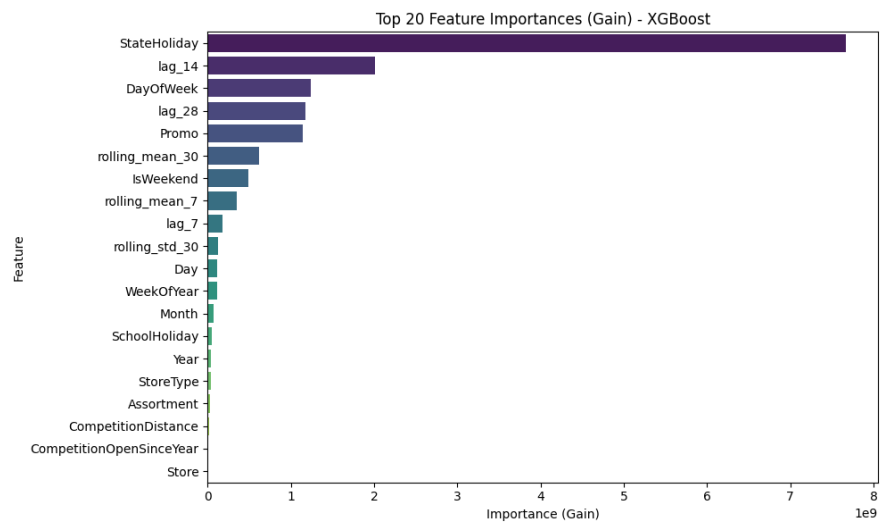
نمودار های تحلیل اکتشافی روی داده های ورودی



مقایسه مقدار واقعی داده ها و مقدار های پیش بینی شده توسط مدل اصلی و مدل بنچمارک



میزان اهمیت ویژگی ها و مقایسه دقت مدل ها با متریک های استاندارد



محدودیت ها و مشکلات فنی فعلی

در نسخه فعلی پروژه، محدودیت هایی از جمله تمرکز روی یک مدل مشخص (XGBoost)، عدم بهره گیری کامل از قابلیت های توضیح پذیری مدل (XAI) و وابستگی به دیتاست خاص Rossmann وجود دارد. همچنین، برخی ویژگی های پیچیده فروشگاه و تعاملات زمانی هنوز به صورت عمومی و قابل تعمیم به دیگر دیتاست ها قالب بندی نشده اند.

برنامه ادامه پروژه

○ برنامه توسعه و بهبود ها تا پایان ترم

- توسعه بخش XAI برای ارائه توضیح دقیق تر درباره تاثیر هر ویژگی بر پیش بینی فروش، با استفاده از SHAP و Partial Dependence Plots.
- ایجاد قالب عمومی مدل و pipeline انعطاف پذیر که بتواند روی سایر دیتاست های فروش یا خرده فروشی با حداقل تغییر اجرا شود.
- بهبود مهندسی ویژگی ها و اضافه کردن قابلیت تعمیم ویژگی های زمانی، پروموشن و رقابت برای مدل های جدید.

○ پیشنهادی برای نسخه نهایی برنامه

نسخه نهایی پروژه شامل یک مدل تعمیم پذیر، سیستم آموزش و اعتبارسنجی انعطاف پذیر و داشبورد XAI تعاملی خواهد بود تا هم پیش بینی دقیق فروش انجام شود و هم تصمیم گیری عملیاتی با شفافیت کامل پشتیبانی گردد.

پتانسیل پروژه

○ جنبه های نوآورانه پروژه

- استفاده از ترکیب ویژگی های زمانی، عملیاتی و فروشگاهی برای تولید پیش بینی های قابل تفسیر و عملیاتی.
- پیاده سازی خط لوله داده ای قابل تکرار که از feature store و اعتبارسنجی زمانی پشتیبانی می کند.
- تمرکز بر تفسیر پذیری مدل (XAI یا SHAP) تا پیش بینی ها قابلیت تبدیل مستقیم به تصمیم های تاملینی و بازاریابی را داشته باشند.

○ امکانات پژوهشی و توسعه ای

- ارزیابی مزایا و محدودیت های مدل های درختی نسبت به معماری های عمیق (LSTM/TFT) روی مقیاس فروشگاه × روز.

- توسعه پیش‌بینی‌های احتمالاتی (quantile forecasting) برای بهینه‌سازی ایمن و سیاست‌های سفارش.
- مطالعه اثر الحاق داده‌های خارجی (هوا، رویداد محلی، رقبا) بر دقت پیش‌بینی در سطوح محلی.
- خودکارسازی بازآموزی مدل مبتنی بر معیار‌های عملکرد و طراحی آزمایش‌های A/B برای ارزیابی اثر تصمیمات پیشنهادی.

منابع

- [The Role of Predictive Analytics in Sales Forecasting](#)
مروری کاربردی بر نقش تحلیل‌های پیش‌بینی‌کننده در بهبود دقت پیش‌بینی فروش و پشتیبانی از تصمیم‌گیری‌های تجاری، با تمرکز بر استفاده از داده‌های تاریخی و مدل‌های تحلیلی در محیط‌های رقابتی (مرجع اصلی)
 - [Time series forecast of sales volume based on XGBoost](#)
بررسی استفاده از XGBoost برای پیش‌بینی حجم فروش با مهندسی ویژگی‌های زمانی و داده‌های تکمیلی.
 - [Forecasting Retail Sales using Machine Learning Models](#)
مطالعه مقایسه‌ای مدل‌های ML شامل رگرسیون خطی، Random Forest، Gradient Boosting، XGBoost و ARIMA برای پیش‌بینی فروش خرده‌فروشی.
 - [Comprehensive Review on Sales Prediction Models](#)
مروری بر مدل‌های مختلف پیش‌بینی فروش خرده‌فروشی از جمله Random Forest، XGBoost و سایر روش‌ها.
 - [Explaining XGBoost predictions with SHAP value: a comprehensive guide to interpreting decision tree-based models](#)
بررسی چارچوب تفسیر نتایج مدل‌های مبتنی بر درخت مانند XGBoost با استفاده از مقادیر SHAP، با تمرکز بر ارزش‌دهی به تاثیر هر ویژگی بر پیش‌بینی‌ها و افزایش شفافیت مدل‌های پیچیده.
 - [Sales Forecasting Using Machine Learning Algorithm in Rossmann Dataset](#)
مطالعه‌ای که داده‌های Rossmann را با مدل‌های مختلف از جمله XGBoost و Prophet تحلیل می‌کند.
 - [XGBoost: A Scalable Tree Boosting System](#)
معرفی الگوریتم XGBoost که پایه مدل ماست (اصلی‌ترین مرجع الگوریتم).
- تمامی روش‌ها، مراحل مهندسی داده، آموزش و اعتبارسنجی مدل، تحلیل عملکرد و نکات اجرایی در مخزن **GitHub** پروژه مستند شده است تا امکان ارزیابی علمی، بازتولید نتایج و توسعه آتی کاملاً مشخص باشد.

مهارت های کسب شده و تقویت شده در این پروژه

○ جست و جوی منابع علمی

- با بسیاری از روش های جست و جوی مقاله و منابع علمی آشنا بودیم، اما در این پروژه یاد گرفتیم چگونه منابع مرتبط تر و کاربردی تر را انتخاب کنیم.
- از ابزار های مبتنی بر AI برای پیدا کردن مقالات مرتبط و درک سریع تر مفاهیم استفاده کردیم، بدون اتکا مستقیم به خروجی آن ها.

○ مهارت نگارش علمی

- ساختاردهی متن (مقدمه، بدنه، نتیجه گیری) که قبلاً با آن آشنا بودیم، در این پروژه منسجم تر و دقیق تر انجام شد.
- توانستیم مهارت خلاصه نویسی مقالات و تبدیل مطالب فنی به متن قابل فهم را بهبود دهیم. باشه، این بخش رو خیلی کوتاه، رسمی و کلی می نویسم، دقیقاً در حد یک فایل پروژه/پیشنهاد و هماهنگ با کدهایی که در نوت بوک پیاده سازی کردیم:
- با بازنویسی مطالب و استفاده درست از AI، از کپی برداری مستقیم پرهیز کردیم.

○ رفرنس نویسی و ارجاع دهی

- آشنایی اولیه با ارجاع دهی داشتیم، که در این پروژه نحوه ی استفاده ی صحیح تر از منابع در متن و فهرست مراجع را تمرین کردیم.
- از AI برای بررسی ساختار ارجاع ها و یکدست سازی آن ها استفاده کردیم.

○ مهارت های نرم افزاری و فنی

- دانش قبلی ما در مباحث علم داده و یادگیری ماشین پایتون تقویت شد و با روش ها و مدل های جدید آشنا شدیم.
- یاد گرفتیم چگونه مدل ها را به شکل منظم تر پیاده سازی و نتایج را بهتر تحلیل کنیم.
- استفاده هدفمند از ابزار های AI به ما کمک کرد فرآیند کار را با سرعت بیشتری انجام دهیم و خطا ها را راحت تر شناسایی و اصلاح کنیم.

○ تفکر سیستمی و حل مسئله

- توانستیم یک مسئله فنی پیش بینی فروش را به یک مسئله تجاری قابل درک تبدیل کنیم.
- نیازمندی های سیستم را مرحله به مرحله شناسایی کردیم و بین داده، مدل و تصمیم تجاری ارتباط برقرار نمودیم.

- این پروژه دید ما را نسبت به کاربرد واقعی تحلیل داده در تصمیم گیری های کسبوکار تقویت کرد.

خود ارزیابی

- درک مسئله اصلی: 5 از 5
- جستجوی منابع علمی: 4 از 5
- نگارش گزارش: 5 از 5
- پیاده سازی فنی: 4.5 از 5
- مدیریت زمان پروژه: 4 از 5