نام و نام خانوادگی : محمد امین فهیمی

حوزه فعالیت: هوش مصنوعی

تاریخ شروع همکاری: 24/4/1404

تاریخ گزارش عملکرد: 28/4/1404

شماره گزارش:3

قسمت اول: فاز مطالعاتی زمان صرف شده : 1 ساعت

1. بررسی scikit-learn و عدم توانایی انجام پروژه توسط scikit-learn : [توضیحات](#توضیحات)
2. [ارزیابی استفاده از Scikit-learn در پروژه Time Series Damage Prediction](#قابلیت_های_پروژه)

امور بروکراسی انجام شده زمان صرف شده: ... ساعت

اولویت‌های کاری فردا

محتوای پیشنهادی جهت کارگذاری در سایت شرکت به نام شما:

* حد اقل یک صفحه تا 5 صفحه باشد
* می تواند دارای عکس باشد
* ترجیحا از دست آورد های شما در زمینه مطالعات و یا تحقیقات انجام شده باشد
* هر گزارش عملکرد یک محتوای جدید داشته باشد
* از کپی و پیست کردن محتوای سایت های دیگر خود داری نمایید
* این قسمت از گزارش الزامی می باشد. لذا در تکمیل آن دقت نمایید
* پیشنهاد: کل گزارش عملکرد می تواند 5 صفحه باشد. لذا از نگارش عجولانه خودداری نمایید و از صبح صفحه نگارش گزارش عملکرد شما باز باشد تا موردی از قلم حذف نگردد و در پایان کار ، مواجهه با حجم زیاد محتوای نگارش نشده گزارش عملکرد نشوید.

**چرا این پروژه با scikit-learn انجام‌پذیر نیست؟**

**📌 ۱. مدل‌های مورد نیاز در پروژه:**

* پروژه نیاز به پیاده‌سازی **LSTM**، **GRU** و **Transformer** دارد.
* اما scikit-learn فقط مدل‌های ساده دارد (مثل رگرسیون، SVM، درخت تصمیم)  
  ✔️ این مدل‌ها **اصلاً پشتیبانی نمی‌شوند** در scikit-learn.

**📌 ۲. پشتیبانی از داده‌های سری زمانی**

* داده‌های سری زمانی نیاز به مدل‌هایی دارند که **ترتیب و وابستگی بین زمان‌ها** را درک کنند.
* scikit-learn فقط با داده‌های **ثابت و بدون ترتیب** کار می‌کند.  
  ✔️ بنابراین برای دنباله‌ها و توالی‌ها مثل سری زمانی مناسب نیست.

**📌 ۳. پیش‌بینی چند خروجی پیچیده (multi-output)**

* پروژه شامل **دو خروجی متفاوت** است: یکی باینری (0/1)، یکی چندکلاسه (مثلاً 3 نوع آسیب)
* scikit-learn نمی‌تواند به راحتی **دو خروجی با loss متفاوت** را همزمان آموزش دهد.  
  ✔️ فقط در شرایط ساده multi-output را پشتیبانی می‌کند، نه در این سطح ترکیبی و پیچیده.

**📌 ۴. نیاز به سفارشی‌سازی پیشرفته**

* پروژه نیاز دارد به:
  + تابع زیان ترکیبی (binary + categorical)
  + early stopping
  + track کردن accuracy و F1 به صورت جداگانه
  + پیاده‌سازی معماری‌های سفارشی
* این کارها فقط در **PyTorch یا TensorFlow** امکان‌پذیر هستند.

**✅ نتیجه‌گیری نهایی:**

❌ این پروژه را **نمی‌توان با scikit-learn انجام داد** چون:

* فاقد پشتیبانی از مدل‌های عمیق (LSTM/GRU/Transformer) است
* داده‌های ترتیبی (سری زمانی) را به درستی مدیریت نمی‌کند
* امکان آموزش همزمان چند خروجی پیچیده را ندارد
* انعطاف لازم برای پیاده‌سازی ساختارهای سفارشی و تنظیم دقیق آموزش را ندارد

✅ برای اجرای کامل این پروژه، باید از فریم‌ورک‌های **PyTorch یا TensorFlow** استفاده کنی.

|  |  |  |
| --- | --- | --- |
| **بخش پروژه** | **اجرا با scikit-learn** | **دلیل** |
| ✅ تولید داده با generator.py | ✅ ممکنه | این مربوط به NumPy/Pandas هست نه یادگیری ماشین |
| 🔍 حذف ویژگی‌های بی‌اثر | ✅ انجام می‌شود | با feature\_importances\_، correlation, یا SelectKBest می‌شود |
| 📊 نرمال‌سازی / استانداردسازی | ✅ انجام می‌شود | با StandardScaler یا MinMaxScaler |
| 🧩 ساخت پنجره‌های زمانی (Sliding Window) | ✅ به صورت دستی | با کد خودت یا تابع آماده، نه خودکار در scikit-learn |
| 🔀 تقسیم‌بندی Train/Val/Test | ✅ انجام می‌شود | با train\_test\_split |
| ❌ ساخت مدل LSTM | ❌ انجام نمی‌شود | مدل LSTM در scikit-learn وجود ندارد |
| ❌ ساخت مدل GRU | ❌ انجام نمی‌شود | GRU هم فقط در PyTorch/Keras هست |
| ❌ ساخت مدل Transformer | ❌ انجام نمی‌شود | scikit-learn هیچ attention model ندارد |
| ❌ پیش‌بینی **همزمان** Active\_Damage و Type\_Damage | ❌ به سختی یا اصلاً نمی‌شود | ترکیب lossها، خروجی‌های چندنوعی (binary + multiclass) ممکن نیست |
| ❌ ترکیب loss‌ها (BCE + CE) | ❌ انجام نمی‌شود | scikit-learn فقط lossهای ساده دارد |
| ❌ سفارشی early stopping | ❌ قابل پیاده‌سازی نیست | باید مدل training loop دستکاری شود |
| ❌ پیاده‌سازی معماری دلخواه | ❌ نمی‌شود | scikit-learn از معماری دلخواه پشتیبانی نمی‌کند |
| ❌ ذخیره وزن‌های مدل‌های عمیق | ❌ ندارد | وزن‌ مدل‌های sklearn ذخیره‌پذیرند ولی فقط برای مدل‌های کلاسیک |
| ❌ تجسم توالی زمانی و attention | ❌ ندارد | Visualization باید در matplotlib/PyTorch ساخته شود |

**🎯 هدف پروژه:**

طراحی و توسعه یک سیستم هوشمند برای **پیش‌بینی هم‌زمان** دو برچسب Active\_Damage و Type\_Damage از داده‌های تولیدشده (Synthetic Time Series)، با استفاده از مدل‌های یادگیری عمیق مانند LSTM، GRU و Transformer.

**✅ قابلیت‌های Scikit-learn در این پروژه:**

Scikit-learn برای مراحل مقدماتی پروژه عملکرد بسیار مناسبی دارد و می‌توان از آن به‌عنوان بخش «Baseline Pipeline» استفاده کرد:

| **توضیح** | **ابزار Scikit-learn** | **مرحله** |
| --- | --- | --- |
| شناسایی nullها، بررسی توزیع، شناسایی outlierها | pandas, sklearn.preprocessing | 1. بارگذاری و تحلیل اولیه داده |
| کاهش ویژگی‌های زائد و افزایش عملکرد مدل | SelectKBest, mutual\_info\_classif, feature\_importances\_ | 2. انتخاب ویژگی‌های مؤثر |
| کدگذاری برچسب‌های Type\_Damage برای مدل‌های ML | LabelEncoder, OneHotEncoder | 3. Encoding برچسب‌ها |
| نرمال‌سازی برای مدل‌های حساس به مقیاس | StandardScaler, MinMaxScaler | 4. مقیاس‌گذاری داده |
| تقسیم به train/test و نگه‌داشتن ترتیب زمانی با shuffle=False | train\_test\_split | 5. تقسیم‌بندی داده‌ها |
| برای ارزیابی اولیه دقت مدل‌های ساده | RandomForestClassifier, LogisticRegression, KNeighborsClassifier | 6. ساخت مدل baseline |
| پیش‌بینی هم‌زمان دو برچسب | MultiOutputClassifier یا MultiOutputRegressor | 7. Multi-output prediction |
| بررسی دقت، یادآوری، F1 و ماتریس خطا | classification\_report, f1\_score, confusion\_matrix | 8. ارزیابی عملکرد |
| بهینه‌سازی hyperparameterها | GridSearchCV, RandomizedSearchCV | 9. تنظیم پارامترها |
| ذخیره و بارگذاری مدل در تولید | joblib | 10. ذخیره مدل نهایی |

**❌ محدودیت‌های Scikit-learn در این پروژه:**

اما برای اجرای کامل پروژه با اهداف یادگیری عمیق، Scikit-learn به دلایل زیر **کفایت نمی‌کند**:

| **مورد** | **محدودیت فنی** |
| --- | --- |
| ❌ عدم پشتیبانی از مدل‌های Sequence-Based | مدل‌هایی مانند LSTM، GRU و Transformer در Scikit-learn پیاده‌سازی نشده‌اند |
| ❌ ناتوانی در ساخت Sliding Windows برای داده زمانی | فاقد ابزارهای آماده برای windowing داده‌های time series |
| ❌ ناتوانی در مدیریت وابستگی زمانی بین رکوردها | الگوریتم‌های Scikit-learn فرض می‌کنند داده‌ها مستقل‌اند |
| ❌ پشتیبانی نکردن از GPU و موازی‌سازی مدرن | زمان آموزش در داده‌های بزرگ ممکن است زیاد باشد |
| ❌ ساختار انعطاف‌پذیر شبکه‌های عمیق | Scikit-learn مدل‌های پیچیده و چند لایه را پشتیبانی نمی‌کند |

**📌 نتیجه‌گیری تخصصی:**

در این پروژه، Scikit-learn تنها برای ساخت یک مدل پایه (Baseline ML Pipeline) کاربرد دارد. جهت طراحی مدل‌های عمیق زمانی (LSTM/GRU/Transformer) که توانایی درک توالی و وابستگی زمانی داده‌ها را دارند، باید از **کتابخانه‌های یادگیری عمیق مانند TensorFlow یا PyTorch** استفاده شود.

**💡 پیشنهاد:**

برای تکمیل پروژه به‌صورت حرفه‌ای، می‌توان دو مسیر موازی ایجاد کرد:

1. **Scikit-learn** به عنوان مرجع مقایسه با **Baseline ML Modeling**
2. **Advanced Deep Sequence Modeling با PyTorch یا TensorFlow**