



دانشگاه صنعتی خواجه نصیرالدین طوسی

دانشگاه صنعتی خواجه نصیرالدین طوسی

final course project

Course: Machine Learning

Student Name: Hadi Fathipour

Student Number: 40411334

خلاصه پروژه

هدف این پروژه، تشخیص آلام آتش‌سوزی از روی داده‌های حسگرهای IoT است. برای این منظور، یک مدل شبکه عصبی چندلایه و یک مدل SVM آموزش داده شده‌اند و عملکرد آن‌ها با معیارهای مختلف مقایسه شده است. در این گزارش، مدل نهایی، نحوه آموزش، نمودارهای اصلی، جدول مقایسه و تحلیل خطا ارائه می‌شود.

معرفی داده‌ها

داده‌ها شامل ویژگی‌های حسگرهای محیطی (دما، رطوبت، eCO_2 ، TVOC، اتانول خام و فشار) و برچسب Fire Alarm هستند. برچسب برابر ۱ نشان‌دهنده رخداد آتش و ۰ نشان‌دهنده حالت عادی است.

پیش‌پردازش

مراحل اصلی پیش‌پردازش به صورت زیر انجام شده است:

- بررسی همبستگی ویژگی‌ها با برچسب Fire Alarm.
- حذف ویژگی‌های کم‌اثر بر اساس تحلیل همبستگی و دانش مسئله.
- شناسایی و حذف داده‌های پرت با روش IQR.
- بررسی مقادیر گم‌شده و توزیع برچسب‌ها.

نکته درباره نشت داده در نسخه اولیه، کران‌های داده پرت از کل داده‌ها محاسبه شده بود. برای ارزیابی دقیق‌تر، یک ارزیابی بدون نشت داده نیز انجام شده که در آن کران‌های پرت تنها از داده‌های آموزش محاسبه و به داده‌های اعتبارسنجی/آزمون اعمال می‌شوند.

مدل نهایی

مدل نهایی یک شبکه عصبی چندلایه (MLP) است که به صورت زیر تعریف شده است:

- دو لایه مخفی با تعداد نرون‌های [40, 20]
- تابع فعال‌سازی سیگموید
- Batch Normalization و Dropout برای کاهش بیش‌برازش
- لایه خروجی با سیگموید برای طبقه‌بندی دودویی

نحوه آموزش

آموزش به دو روش انجام شده است:

۱. روش **Holdout**: تقسیم داده به آموزش و آزمون و گزارش معیارها.
۲. اعتبارسنجی **K-Fold**: آموزش و ارزیابی در چند تکرار برای کاهش بایاس ارزیابی.

نمودارها

نمودارهای زیر برای تحلیل مدل و داده‌ها استفاده شده‌اند:

- نمودار همبستگی ویژگی‌ها با برچسب
- نمودار جعبه‌ای برای مشاهده داده‌های پرت
- نمودار تاریخچه آموزش (کاهش خطا و تغییر معیارها)
- Confusion Matrix
- منحنی ROC
- نمودار SHAP برای تفسیر ویژگی‌ها

جایگذاری نمودارها (خروجی‌های نوت‌بوک)

شکل ۱: نمونه محل قرارگیری نمودارها

جدول مقایسه مدل‌ها

در جدول زیر معیارهای اصلی مقایسه شده‌اند. اعداد باید پس از اجرای نهایی نوت‌بوک قرار داده شوند.

مدل	Accuracy	Precision	Recall	F1	AUC	Loss
شبکه عصبی	□	□	□	□	□	□
SVM	□	□	□	□	□	□

تحلیل خطا

تحلیل خطا بر اساس Confusion Matrix انجام می‌شود. موارد مهم:

□ **False Negative**: آتش واقعی که مدل آن را تشخیص نداده است (ریسک بالا).

□ **False Positive**: هشدار اشتباه (هزینه عملیاتی).

بررسی می‌شود که آیا مدل به سمت یکی از کلاس‌ها سوگیری دارد یا خیر، و آیا آستانه تصمیم‌گیری باید تنظیم شود.

نتیجه‌گیری

در این پروژه، مدل شبکه عصبی عملکرد بالایی در تشخیص آلام نشان داد و با مدل SVM مقایسه شد. با توجه به نتایج، شبکه عصبی عملکرد بهتری در معیارهای کلی داشت، اما تحلیل خطا نشان می‌دهد که باید روی کاهش False Negative تمرکز ویژه داشت. همچنین اجرای ارزیابی بدون نشت داده، تصویر دقیق‌تری از تعمیم‌پذیری مدل ارائه می‌دهد.