

## هدف برنامه:

هدف این برنامه طراحی و پیاده‌سازی یک مدل شبکه عصبی LSTM برای شناسایی و دسته‌بندی داده‌ها در یک مجموعه داده‌ی بزرگ است. داده‌ها به کمک LSTM پردازش می‌شوند تا روابط زمانی بین ویژگی‌ها تشخیص داده شود و دقت مدل در پیش‌بینی به حداکثر برسد.

## پیش‌پردازش داده‌ها:

1- بارگذاری و آماده‌سازی داده‌ها: داده‌ها از یک فایل CSV بارگذاری شده و نام ستون‌ها تنظیم شده است.

2- پاکسازی داده‌ها: مقادیر 'inf' به 'NaN' تبدیل شده و داده‌های خالی حذف شدند.

3- کدگذاری برچسب‌ها: برچسب‌ها از حالت متنی به اعداد تبدیل شدند تا شبکه عصبی قابل درک باشد.

4- مقیاس‌بندی ویژگی‌ها: داده‌ها به کمک 'StandardScaler' نرمال‌سازی شدند تا همه ویژگی‌ها در یک مقیاس مشابه قرار گیرند و مدل عملکرد بهتری داشته باشد.

5- تقسیم داده‌ها: داده‌ها به دو مجموعه آموزش و آزمایش با نسبت ۷۰ به ۳۰ تقسیم شدند.

6- ساختار زمانی: داده‌ها به یک ساختار زمانی (توالی داده‌ها با طول ۱۰۰) تبدیل شدند تا مدل بتواند روابط زمانی را بیاموزد.

## ساخت و آموزش مدل: LSTM

یک مدل ساده LSTM با لایه‌های زیر طراحی و آموزش داده شد:

لایه LSTM: با ۶۴ واحد که وظیفه استخراج روابط زمانی را دارد.

لایه Dropout: برای جلوگیری از بیش‌برازش و بهبود تعمیم‌دهی مدل.

لایه Dense: یک خروجی با تابع فعال‌سازی 'sigmoid' برای پیش‌بینی کلاس‌ها.

## آموزش مدل:

مدل به مدت ۱۰ دوره (Epoch) با استفاده از 'Adam Optimizer' و 'Binary Crossentropy Loss' آموزش دید. در هر دوره، مدل دقت (Accuracy) خود را بهبود داده و میزان خطا (Loss) کاهش یافت.

## نتایج و ارزیابی مدل:

نتایج نهایی مدل بر روی داده‌های آزمایشی به شرح زیر است:

**Accuracy**(دقت): ۰.۹۹۹، یعنی مدل توانسته ۹۹.۹ درصد از نمونه‌ها را به درستی پیش‌بینی کند.

**Precision**(دقت مثبت) و **Recall** (بازخوانی) برای هر دو کلاس نزدیک به ۱.۰۰ هستند، که نشان‌دهنده عملکرد بسیار بالای مدل در شناسایی درست کلاس‌ها است.

**F1 Score** : مقدار ۰.۹۹۹۳، که به معنای تعادل بسیار خوب بین **Precision** و **Recall** است.

**ROC AUC Score** : ۰.۹۹۹، که نشان‌دهنده توانایی بسیار بالای مدل در تمایز بین کلاس‌ها می‌باشد.

نتیجه‌گیری:

مدل **LSTM** پیاده‌سازی شده توانسته است با دقت و عملکرد بالایی داده‌های آزمایشی را پیش‌بینی کند. دقت و امتیازات ارزیابی (**F1 Score** و **ROC AUC**) نشان می‌دهند که مدل به خوبی یاد گرفته و عملکرد بالایی دارد. با این حال، برای داده‌های واقعی و در سناریوهای کاربردی، ارزیابی بیشتر و شاید بهینه‌سازی بیشتری نیاز باشد تا تعمیم‌دهی مدل بهبود یابد.

نمودارها:

تغییرات دقت مدل در طی ۱۰ دوره به کمک نمودار رسم شده است که نشان‌دهنده افزایش دقت و کاهش خطا در طی زمان آموزش است.

پیشنهادهای:

1- افزایش تعداد داده‌های آموزش می‌تواند مدل را بیش از پیش بهبود بخشد.

2- استفاده از شبکه‌های عصبی پیچیده‌تر مانند **Bi-LSTM** یا افزودن ویژگی‌های دیگری می‌تواند منجر به افزایش دقت و تعمیم‌دهی مدل شود.