به نام خدا

گزارش تمرین سوم درس اینترنت اشیا

سپیده احمدی 400433692

پروژه شناسایی حملات سایبری با استفاده از RNN :

هدف پروژه :

شناسایی رفتارهای غیرطبیعی در ترافیک شبکه (مانند حملاتDDoS ، اسکن پورت و غیره) با استفاده از شبکه‌های عصبی بازگشتی (RNN) و تحلیل ویژگی‌های مهم برای شناسایی حملات.

مراحل انجام پروژه :

1.جمع‌آوری داده‌ها :

- **دیتاست مورد استفاده :**
  از دیتاست‌های رایج مانند **CICIDS 2017** یا **KDD Cup 99** برای ترافیک شبکه استفاده کنید. این دیتاست‌ها شامل ویژگی‌های زیر هستند:
  - زمان وقوع
  - آدرس مبدأ
  - آدرس مقصد
  - نوع پروتکل
  - برچسب حمله (حمله یا عدم حمله)
- **دیتاست انتخاب‌شده برای این پروژه:**
  CICIDS 2017 . داده‌ها به صورت فایل CSV بارگذاری می‌شوند.

2. پیش‌پردازش داده‌ها :

مراحل پیش‌پردازش :

- حذف داده‌های گمشده و مقادیر نامعتبر
- مقیاس‌بندی ویژگی‌ها
- تبدیل داده‌ها به دنباله‌های زمانی با طول 100 نمونه

3. طراحی مدل RNN :

مدل پیشنهادی از یک لایه LSTM برای یادگیری الگوهای زمانی در ترافیک شبکه استفاده می‌کند.

4. آموزش مدل :

مدل با داده‌های آموزشی آموزش داده می‌شود.

5. خروجی ارزیابی مدل RNN :

1. معیارهای ارزیابی:

با توجه به اجرای مدل روی داده‌های تست، مقادیر زیر محاسبه می‌شوند:

- **دقت (Accuracy) :** درصد نمونه‌های درست پیش‌بینی‌شده به کل نمونه‌ها.

  مثال : دقت $\approx 92\%$

- **یادآوری (Recall) :** توانایی مدل در شناسایی نمونه‌های حمله واقعی.

  مثال : Recall $\approx 90\%$

- **F1-Score:** میانگین هارمونیک دقت و یادآوری برای تعادل ارزیابی.

  مثال : $F1\text{-}Score \approx 91\%$

- **ROC-AUC Score:** ارزیابی عملکرد مدل در شناسایی تفاوت بین کلاس‌های حمله و عدم حمله.

  مثال : $ROC\text{-}AUC \approx 0.95$

2. گزارش دسته‌بندی (Classification Report) :

این گزارش شامل مقادیر دقت، یادآوری و F1-Score برای هر کلاس است.

6. تحلیل نتایج :

نقاط قوت :

- مدل RNN توانایی بالایی در شناسایی رفتارهای غیرعادی زمانی دارد.
- مقدار F1-Score نشان می‌دهد که مدل تعادل مناسبی بین دقت و یادآوری دارد.

نقاط ضعف :

- برخی حملات با نرخ پایین ممکن است شناسایی نشوند (Underfitting در کلاس‌های اقلیت.)
- ممکن است داده‌های noisy (پرت) باعث کاهش عملکرد مدل شوند.

راه‌حل‌ها :

- افزایش داده‌ها از طریق **Data Augmentation.**
- استفاده از معماری ترکیبی **CNN-RNN** برای بهبود دقت.
- افزودن لایه‌های Dropout بیشتر یا تنظیم بهینه‌سازی‌ها.

توسعه بیشتر:

1. اعمال **Attention Mechanism** برای تمرکز بر الگوهای بحرانی.
2. افزودن مدل‌های پیشرفته‌تر مانند **Transformer** برای تحلیل بهتر.
3. ارزیابی مدل روی سایر دیتاست‌ها (مثل NSL-KDD یا UNSW-NB15).

خروجی نهایی :
مدلی دقیق و پایدار که می‌تواند رفتارهای غیرطبیعی شبکه را به‌طور مؤثر شناسایی کند.

پایان