

Nama: M.Rizky Fadillah (1103210259)

Kelas: TK-45-GAB

## LAPORAN TUGAS 14 PEMBELAJARAN MESIN

### - **Bidirectional RNN Model**

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model **Bidirectional Recurrent Neural Network (BiRNN)** dalam mendeteksi URL phishing. Dataset yang digunakan adalah **PhiUSIIL\_Phishing\_URL\_Dataset**, yang berisi fitur-fitur numerik yang menggambarkan URL, dengan target yang menunjukkan apakah URL tersebut phishing atau bukan.

Eksperimen dilakukan untuk menguji berbagai kombinasi parameter yang memengaruhi kinerja model, termasuk ukuran lapisan tersembunyi, metode pooling, jumlah epoch, dan pilihan optimizers.

### **2. Metodologi**

#### **2.1. Persiapan Data**

Dataset yang digunakan berisi data URL yang telah diproses dan diberi label sebagai phishing atau tidak. Proses pra-pemrosesan data dilakukan dengan langkah-langkah berikut:

1. **Pemuatan Dataset:** Dataset dimuat menggunakan Pandas, kemudian diambil sampel sebanyak 10.000 data untuk eksperimen.
2. **Seleksi Fitur dan Target:** Fitur yang digunakan adalah variabel numerik dalam dataset, dan target yang diprediksi adalah kolom label yang menunjukkan apakah URL phishing (1) atau tidak (0).
3. **Pembagian Data:** Dataset dibagi menjadi data pelatihan (80%) dan data pengujian (20%) menggunakan `train_test_split`.
4. **Standarisasi Fitur:** Fitur distandarisasi menggunakan `StandardScaler` untuk mengoptimalkan performa model.
5. **Konversi ke Tensor:** Data pelatihan dan pengujian kemudian dikonversi menjadi tensor PyTorch agar bisa diproses oleh model.

#### **2.2. Model**

Model yang digunakan adalah **Bidirectional Recurrent Neural Network (BiRNN)**. RNN digunakan untuk menangani data urutan, dan pada model ini menggunakan dua arah (bidirectional) untuk memanfaatkan informasi baik dari masa lalu maupun masa depan. Model ini juga dilengkapi dengan dua jenis teknik pooling:

1. **Max Pooling:** Mengambil nilai maksimum dari setiap dimensi output RNN.
  2. **Average Pooling:** Mengambil rata-rata dari setiap dimensi output RNN.
- Output RNN kemudian diteruskan ke layer fully connected untuk menghasilkan prediksi.

#### **2.3. Optimasi dan Evaluasi**

Untuk pelatihan, beberapa optimizer yang umum digunakan diuji, termasuk **SGD**, **RMSPprop**, dan **Adam**. Proses pelatihan menggunakan **Cross Entropy Loss** untuk mengukur performa model dalam klasifikasi. Selain itu, **early stopping** diterapkan untuk menghentikan pelatihan lebih awal jika model tidak menunjukkan perbaikan dalam loss pengujian setelah beberapa epoch.

Evaluasi dilakukan menggunakan dua metrik utama:

- **Loss:** Mengukur seberapa baik model memprediksi kelas.
- **Akurasi:** Menghitung persentase prediksi yang benar pada dataset pengujian.

#### **2.4. Eksperimen**

Berikut adalah eksperimen yang dilakukan untuk mengeksplorasi pengaruh berbagai parameter terhadap performa model:

1. **Ukuran Lapisan Tersembunyi:** Menguji berbagai ukuran lapisan tersembunyi (32, 64, 128).
2. **Metode Pooling:** Menguji dua jenis pooling yaitu **Max Pooling** dan **Average Pooling**.
3. **Jumlah Epoch:** Menguji jumlah epoch yang berbeda (5, 50, 100, 250, 350).
4. **Optimizers:** Menguji tiga optimizer yang berbeda (SGD, RMSProp, dan Adam).

### 3. Hasil Eksperimen

#### 3.1. Eksperimen dengan Ukuran Lapisan Tersembunyi

Hidden Size	Pooling	Optimizer	Epochs	Test Loss	Accuracy
32	max	Adam	50	2.272466e-06	100.00%
64	max	Adam	50	2.358092e-06	100.00%
128	max	Adam	50	1.396050e-07	100.00%

Hasil eksperimen menunjukkan bahwa perubahan ukuran lapisan tersembunyi dari 32, 64, hingga 128 tidak mengubah hasil secara signifikan. Semua eksperimen memberikan akurasi 100% dan loss yang sangat kecil.

#### 3.2. Eksperimen dengan Metode Pooling

Hidden Size	Pooling	Optimizer	Epochs	Test Loss	Accuracy
64	max	Adam	50	7.528792e-04	99.95%
64	avg	Adam	50	2.597874e-06	100.00%

Pada eksperimen ini, hasil pooling menggunakan metode **max pooling** sedikit lebih rendah akurasinya (99.95%) dibandingkan dengan **average pooling** yang mencapai 100% akurasi.

#### 3.3. Eksperimen dengan Jumlah Epoch

Hidden Size	Pooling	Optimizer	Epochs	Test Loss	Accuracy
64	max	Adam	5	3.062082e-06	100.00%
64	max	Adam	50	1.386664e-06	100.00%
64	max	Adam	100	1.020905e-05	100.00%
64	max	Adam	250	9.020944e-07	100.00%
64	max	Adam	350	1.181789e-06	100.00%

Pada eksperimen ini, meskipun jumlah epoch bervariasi antara 5 hingga 350, performa model tetap konsisten dengan akurasi 100% dan loss yang sangat rendah.

### 3.4. Eksperimen dengan Optimizer

Hidden Size	Pooling	Optimizer	Epochs	Test Loss	Accuracy
64	max	SGD	50	2.633669e-03	100.00%
64	max	RMSProp	50	4.020933e-09	100.00%
64	max	Adam	50	1.273837e-06	100.00%

Pada eksperimen ini, ketiga optimizer memberikan hasil yang sangat baik dengan akurasi 100%. Namun, **RMSProp** menghasilkan loss yang lebih kecil dibandingkan dengan **SGD** dan **Adam**.

### 4. Kesimpulan

Berdasarkan hasil eksperimen yang dilakukan, dapat disimpulkan bahwa:

- **Ukuran Lapisan Tersembunyi:** Ukuran lapisan tersembunyi tidak memengaruhi akurasi secara signifikan selama model menggunakan pooling yang tepat dan optimizer yang efektif.
- **Metode Pooling:** Metode **average pooling** memberikan sedikit peningkatan dalam akurasi dibandingkan dengan **max pooling**, meskipun perbedaannya kecil.
- **Jumlah Epoch:** Jumlah epoch yang lebih banyak tidak meningkatkan akurasi lebih jauh, menunjukkan bahwa model telah mencapai konvergensi pada jumlah epoch yang lebih rendah.
- **Optimizers:** Semua optimizer (SGD, RMSProp, dan Adam) bekerja sangat baik, namun **RMSProp** sedikit lebih unggul dalam hal loss.

Secara keseluruhan, model BiRNN yang diterapkan berhasil dengan sangat baik pada tugas deteksi phishing URL ini, dengan akurasi 100% di semua eksperimen.