Raport projektu: Filtrowanie spamu oparte na modelu językowym

1. Wyniki Testowe i Treningowe

Do oceny skuteczności modeli wykorzystano trzy metryki: stratę (loss), dokładność (accuracy) oraz krzywą uczenia. Spośród trzech testowanych architektur (RNN, GRU, LSTM), najlepsze wyniki osiągnął model oparty na warstwach GRU.

Model	Dokładność walidacyjna	Strata walidacyjna
RNN	98.4%	0.06
GRU	98.9%	0.05
LSTM	99.0%	0.07

Na danych testowych model GRU osiągnął dokładność 98.9% oraz stratę 0.05, co świadczy o jego wysokiej skuteczności i zdolności do generalizacji, dzięki czemu uzyskał najlepsze wyniki spośród porównywanych modeli.

2. Uzasadnienie Wyboru Techniki/Modelu

Zdecydowałem się porównać trzy popularne architektury sieci rekurencyjnych, aby sprawdzić, która najlepiej radzi sobie z detekcją spamu. Ostatecznie wybrałem **GRU** (Gated Recurrent Unit), ponieważ oferuje kompromis pomiędzy jakością predykcji a czasem uczenia. GRU zachowuje informacje o kontekście lepiej niż klasyczne RNN, a jednocześnie jest mniej zasobożerny niż LSTM.

3. Strategia Podziału Danych

Zbiór danych został podzielony na zbiór treningowy i walidacyjny przy użyciu funkcji `train_test_split` z biblioteki `scikit-learn`. Zbiór walidacyjny zawiera dokładnie 10 000 przykładów, a pozostała część danych została wykorzystana do treningu modelu. Podział został przeprowadzony w sposób stratyfikowany względem etykiet (parametr `stratify=df['label']`), co zapewnia zachowanie proporcji klas w obu zbiorach. Ustawienie ziarna losowości (`random_state=42`) gwarantuje powtarzalność eksperymentów. Taki podział pozwala na wiarygodną ocenę jakości modelu bez konieczności stosowania oddzielnego zbioru testowego, ponieważ nie były dostrajane hiperparametry.

4. Opis Danych Wejściowych

Dane zostały pobrane z Kaggle – Email Spam Classification Dataset. Zawierają one dwie kolumny:

- **text**: treść wiadomości e-mail
- **label**: etykieta binarna (0 - nie spam, 1 - spam)

Dane zostały poddane minimalnemu przetwarzaniu — utworzono słownik wyrazów (tokenizacja), a następnie zakodowano wiadomości jako ciągi identyfikatorów słów.

5. Analiza Wyników i Propozycje Dalszych Kroków

Model GRU osiągnął wysoką dokładność przy relatywnie niskiej stracie, co sugeruje skuteczne rozpoznawanie spamu. Możliwe ulepszenia obejmują:

- Rozszerzenie słownika i zastosowanie wstępnie wytrenowanych wektorów słów (np. GloVe),
- Eksperymenty z długością sekwencji i głębokością sieci,
- Dodanie prostych reguł heurystycznych jako dodatkowe cechy (np. obecność linków),
- Budowa wersji webowej / REST API do użycia modelu poza aplikacją lokalną.