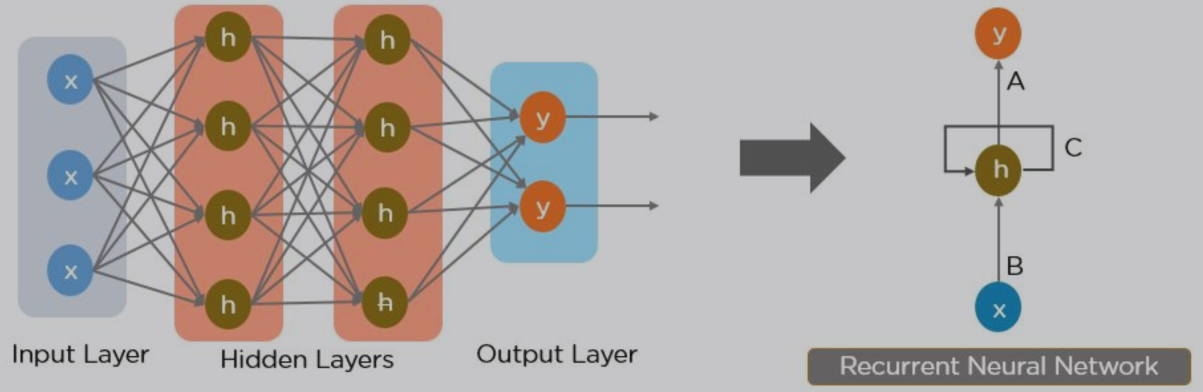
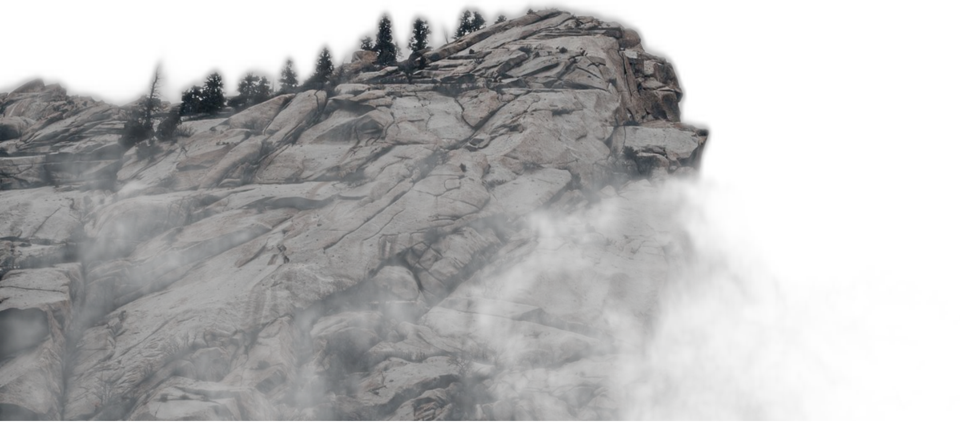


COMPARATIVE ANALYSIS OF DEEP LEARNING MODELS FOR SENTIMENT CLASSIFICATION

RICARDO URBINA, KEVIN NIETO, JOSEPH VELEZ

UNIVERSIDAD ICESI - COMPUTACION Y ESTRUCTURAS DISCRETAS III

Índice



01. Introduction

Sentiment Analysis is a critical NLP task for automating feedback processing. This project conducts a comparative analysis of three neural architectures—Dense Networks (DNN), Vanilla RNNs, and LSTMs—to determine the most effective approach for binary sentiment classification on short-text datasets (Amazon, IMDb, Yelp).

02. Objectives

General

To evaluate and compare the performance of non-sequential vs. sequential deep learning architectures in binary sentiment classification.

02. Objectives

Specific

1.

Implement a robust NLP pipeline (POS-tagging, Lemmatization) to reduce feature sparsity.

2.

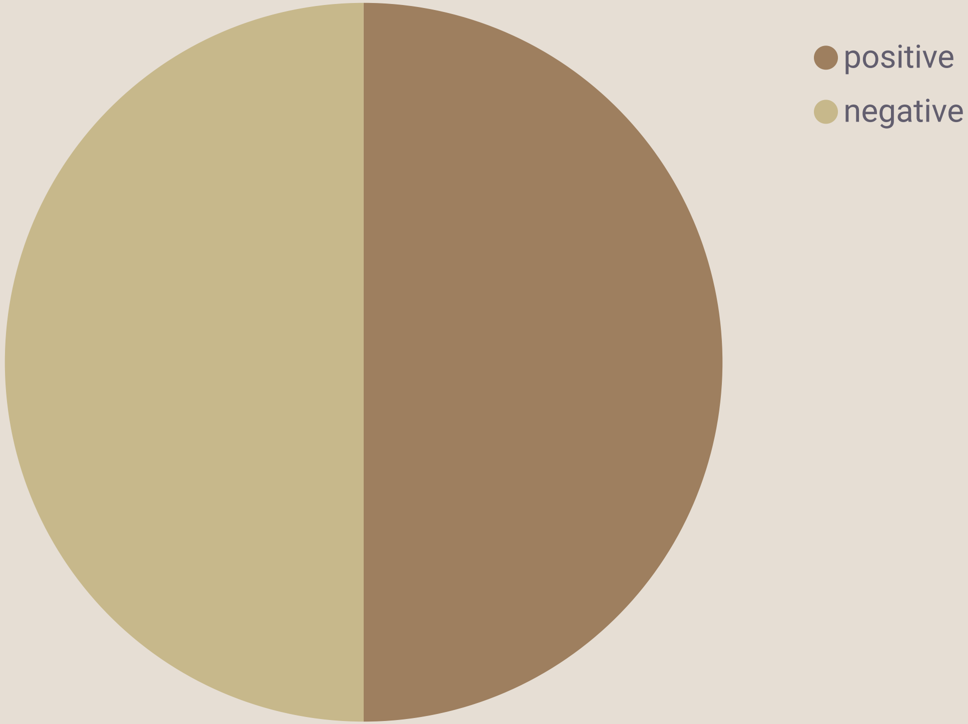
Design and train three distinct models: DNN (TF-IDF), Vanilla RNN, and LSTM.

3.

Apply GridSearchCV with Stratified K-Fold Cross-Validation to tune hyperparameters and control overfitting.

4.

Assess performance using Accuracy, F1-Score, and Cohen's Kappa to identify the optimal architecture for small datasets.



Objective: Binary Classification (Positive vs. Negative).
Volume: 2,726 instances (post-cleaning).
Challenge: Short text length & limited sample size.

Dataset: Kotzias et al. (2015).

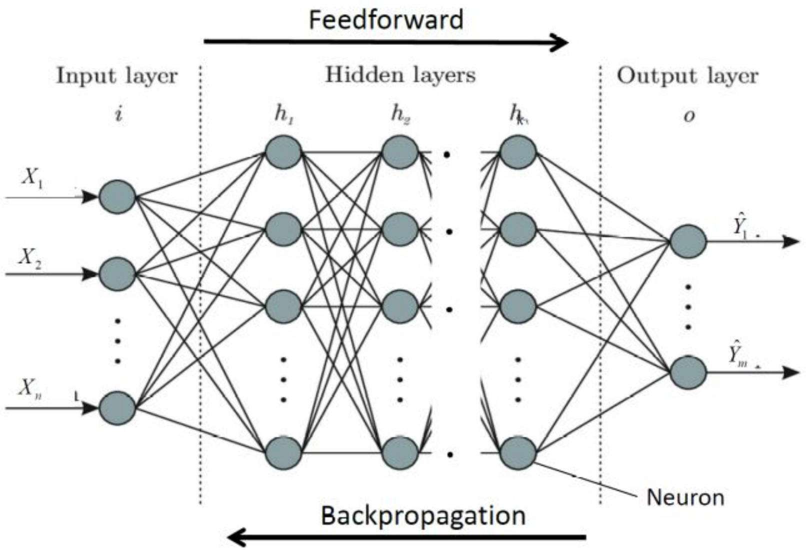
03. Preprocessing Pipeline

1. Normalization
2. Noise Removal
3. Tokenization
4. Lemmatization
5. Stopword Removal

04. Question of Interest

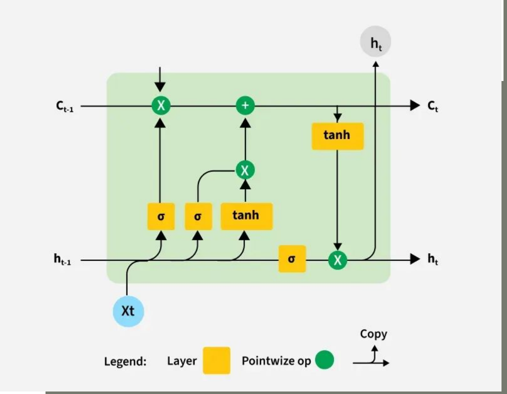
Does the complexity of sequential architectures (RNN, LSTM) yield better performance on short, informal text than a simpler non-sequential approach (DNN with TF-IDF)?

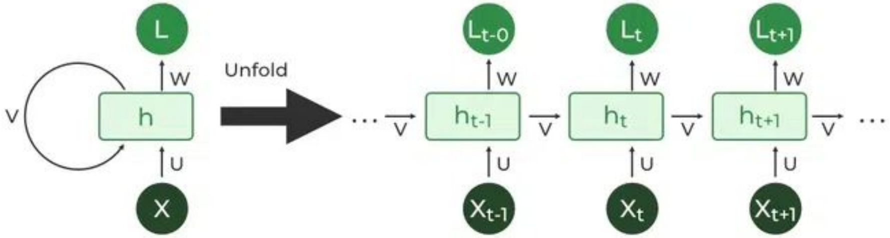
05. DNN



Dense Neural Network (TF-IDF)

- **Input Representation:** TF-IDF (Bag of Words).
- **Vocab Size:** 5,000 features.
- **Architecture:**
  - Hidden Layer 1: 64 units (ReLU) + Dropout.
  - Hidden Layer 2: 32 units (ReLU) + Dropout.
  - Output: Sigmoid.





05. RNN vs. LSTM

Input Representation: Padding Sequences (Length: 11 for RNN, 14 for LSTM).

Embedding Layer: Dim 50 + SpatialDropout1D.

Vanilla RNN: 32 units. Prone to vanishing gradients.

LSTM: 64 units. Handles long-term dependencies via Forget/Input gates.

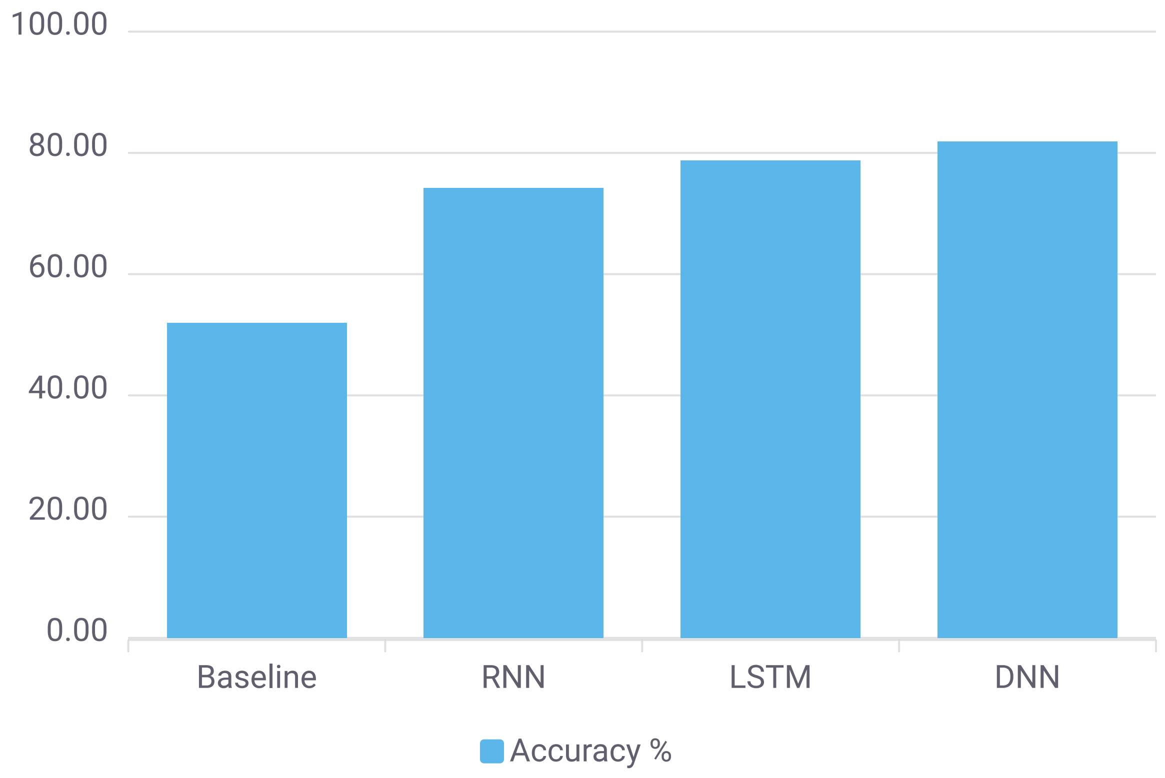
05. Setup & Tuning

Reconstrucción usando distintos valores de *k*

- **Strategy:** GridSearchCV with 3-Fold Cross-Validation.
- **Split:** 80% Train / 20% Test (Stratified).
- **Key Hyperparameters Tuned:**
    - Batch Size (32 vs 64).
    - Optimizer (Adam vs RMSprop).
    - Dropout Rate (0.3 to 0.6).

06. RESULTS

06. Quantitative Results



06. Why did the DNN win?

1

2

3

4

Feature Analysis: Top predictors include "Love", "Great", "Bad", "Poor".

Insight 1: Reviews are very short.

Insight 2: Sequence/Word order matters less than the presence of strong sentiment keywords.

Insight 3: Overfitting was observed in LSTM/RNN due to small data volume.

Con k=40, el ratio sigue excelente (≈6.4) pero con calidad aceptable (PSNR 35.6 dB, SSIM 0.889).

Paso 2: Consolidación de Datos

- Fusiona métricas de compresión y energía con la tabla de calidad (PSNR/SSIM)
- Genera DataFrame completo df_all con todas las métricas por imagen y valor k
- Exporta a calidad_compresion_energia.csv para análisis posterior

Paso 1: Cálculo de Compresión

- Cuantifica el ahorro de almacenamiento comparando:
- Almacenamiento Original: $m \times n$ (escala de grises) o $3 \times m \times n$ (RGB)
- Almacenamiento SVD: $k \times (m+n+1)$ por canal
- Ratio de Compresión = Original/SVD (ej: ratio=50 → 50× menos espacio)

Paso 3: Gráficos de Análisis

Visualiza las relaciones clave mediante gráficos de dispersión:

- Compresión vs. k: Muestra cómo disminuye el ratio al aumentar k
- PSNR vs. Compresión y SSIM vs. Compresión: Evalúan el trade-off calidad/costo
- Identifican el punto de rendimiento decreciente

Paso 4: Resumen de k Óptimos

- Tabla automática que identifica el k mínimo para cumplir objetivos:
- Energía acumulada $\geqslant$ 95%
- SSIM $\geqslant$ 0.95
- PSNR $\geqslant$ 30 dB

Responde directamente nuestra pregunta mostrando que los criterios de calidad no siempre coinciden:

- Para 4.2.01.tiff: k=5 logra >95% energía, k=20 para >30dB PSNR, pero k=120 para >0.95 SSIM (métrica más estricta)
- Imágenes complejas como 4.2.03.tiff (mandril) nunca alcanzan umbrales PSNR/SSIM, siendo "más difíciles" de comprimirReintentarClaude puede cometer errores. Verifique las respuestas

Con k=40, el ratio sigue excelente (≈6.4) pero con calidad aceptable (PSNR 35.6 dB, SSIM 0.889).

Para 4.2.01.tiff con k=5, obtenemos ratio de compresión masivo (≈51.15), pero calidad pobre: PSNR 24.2 dB y SSIM 0.776.

07. Conclusions & Future Work

TF-IDF + DNN is the most efficient model for this specific dataset.

LSTM Capabilities: Showed higher recall (0.84), useful if missing positive reviews is costly.

Future Work:

Pre-trained Embeddings (GloVe).

Data Augmentation to reduce overfitting.

Transformer Models (BERT) for deeper context.

Los resultados experimentales mostraron que una cantidad relativamente pequeña de componentes k (entre 40 y 70 en promedio) ofrece un equilibrio óptimo entre calidad y compresión, manteniendo alta similitud estructural (SSIM) y errores bajos (MSE).

Thanks for reading