COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI ICADE CIHS

Final Proyect Proposal
Gossip Alert Generation

Bishops

Sofía Negueruela Avellaneda
Enrique Fernández-Baíllo Rodríguez de Tembleque
Jacobo Banús Bertram
Almudena Garrido García-Pita

Aprendizaje Profundo &
Procesamiento de Lenguaje Natural
3º Grado en Ingeniería Matemática e Inteligencia Artificial

Proposal

The rise of entertainment journalism and celebrity culture has led to an increasing demand for automatized analysis of news. Our goal is to develop a deep learning model that processes news and articles of this nature and extracts meaningful insights, focusing on Named Entity Recognition (NER) and Sentiment Analysis (SA). Our goal is to process both the text and the pictures of the articles, but we will start with a small part of this challenge at first (text processing) and then implement the harder part (image processing)

Our approach consists of three main steps:

1. **Named Entity Recognition (NER):** For identifying key figures, locations, and organizations within news. For example, in the headline “Celebrity X spotted with a mysterious partner in Paris”, the model should recognize Celebrity X as a person and Paris as a location.

2. **Sentiment Analysis (SA):** To determine whether an article portrays a celebrity positively or negatively. This will help assess media bias or detect trends in public perception. For instance, the headline “Celebrity Y accused of tax fraud” should be classified as negative.

3. **Automatic Summary Generation**: Using NER and SA outputs, a text generation model will produce short, meaningful summaries. These will capture key events, trends, or controversies related to public figures.

Our plan includes:

- Dataset selection or creation using a generative AI.
- Incremental model development and testing, following the three phases proposed in the formulation of the project.
- Final integration and performance evaluation after each phase.
- When this proposal is approved, we will refine our methodology and set clear and more accurate milestones.

First Ideas

For the moment, we think we will use an API to extract news articles. Still, this type of articles follow a particularly clear pattern in terms of sentiment related words (i.e. infidelities or dramas). Regarding the models, we are thinking about using a combination of convolutional and recurrent networks for the image caption generation, and bidirectional LSTMs for text analysis. This type of networks are particularly efficient in NER, since entities might depend on both the past and the future context. We might also combine this model with a Conditional Random Field Layer to improve performance.