

# Stock Market Trend Prediction Using News Sentiment Analysis

---

## Problem Statement

Financial markets are highly sensitive to news. Timely and accurate interpretation of financial news can be a powerful signal for traders, stock analysts, and investors. This project aims to analyze news articles related to stock markets or specific companies and predict a sentiment score that map to actionable investment advice: Buy, Hold, or Sell.

## Objective

To build a deep learning-based sentiment classifier that:

1. Takes in textual financial news headlines
2. Processes the text using NLP and deep learning
3. Outputs a sentiment score between 0 and 1
4. Interprets the score into Buy, Hold, or Sell recommendations

## Why Hybrid Model

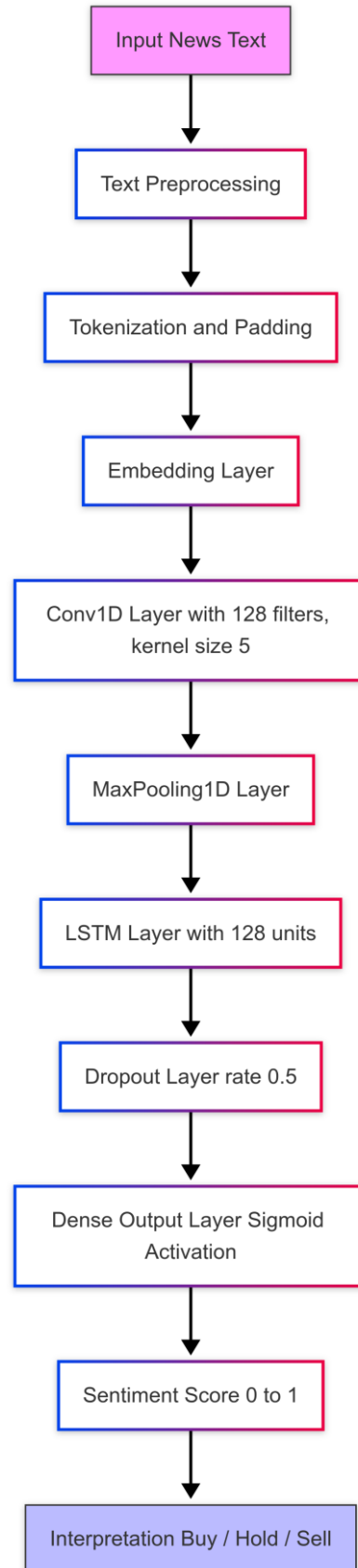
1. CNNs are excellent at capturing short patterns like 'revenue fall' or 'merger announced', which are strong sentiment indicators.
2. LSTMs provide memory and understand the context of the sequence — important when the sentiment depends on the whole sentence (e.g., 'despite profits, growth is slowing').
3. Combining both allows the model to learn semantic patterns and contextual dependencies effectively.
4. The sigmoid output (which is the activation function) gives a fine-grained sentiment score rather than binary output, which suits the stock decision mapping (Buy/Hold/Sell).

## Model Architecture

The model is a hybrid of CNN and LSTM layers designed to capture both local n-gram patterns and long-term dependencies in news headlines.

### Layers Used:

1. **Input Layer** - Raw text headline from financial news (stock market related for any company)
2. **Text Preprocessing** - Lowercasing, punctuation removal, stop words filtering
3. **Tokenization & Padding** - Converts text to sequences of word indices and pads sequences to fixed length
4. **Embedding Layer** - Converts word indices to dense vectors to capture semantic meaning of words
5. **1D Convolutional Layer** - 128 filters, kernel size = 5; captures local n-gram features
6. **MaxPooling1D Layer** - Reduces dimensionality while preserving important features
7. **LSTM Layer** - 128 units; captures long-term dependencies and contextual flow in sentences
8. **Dropout Layer** - Rate = 0.5 to prevent overfitting
9. **Dense Output Layer** - Sigmoid activation; outputs a score between 0 and 1
10. **Interpretation Layer** - Threshold-based conversion:
  - **Score  $\geq 0.6$**  → Buy
  - **Score between 0.4 and 0.6** → Hold
  - **Score  $\leq 0.4$**  → Sell



## **Benefits**

1. Works on historical news
2. Adaptable to mutual funds, SIPs, futures & options.
3. Lightweight enough to be deployed in dashboards or APIs