Executive Summary

Project Title: Impact of Donald Trump Social Media Communication on Financial Market Volatility

Analyst: Dario Meacci
Date: January 2026

1. Main Objective of Analysis

The primary objective of this analysis is to determine if unstructured executive communication, specifically, the social media activity of U.S. President Donald Trump, can serve as a quantifiable leading indicator for financial market volatility (VIX). By leveraging Deep Learning and Natural Language Processing (NLP), this project aims to build a predictive model that identifies the relationship between specific political "sentiment topics" and subsequent market fear, providing a strategic edge for risk management.

2. Data Description

The analysis utilised two primary datasets covering the 2017–2021 presidential term:

- The Input (X): A corpus of 33,669 tweets from Donald Trump, cleaned and preprocessed. We engineered features to move beyond simple volume metrics, categorising text into semantic clusters: *Economy, Geopolitics, Domestic Policy, Corporate Attacks,* and *Other*.
- The Target (y): Historical data from the CBOE Volatility Index (VIX). The problem was framed as a Binary Classification task, where the model predicts whether the VIX will close "Higher" or "Lower" on the following trading day based on the previous day's tweets.

3. Deep Learning Model Variations & Selection

To identify the optimal predictive architecture, we designed and rigorously tested four distinct Deep Learning variations:

- Model 1 (Baseline Dense Network): A simple feed-forward network using topic clusters. (Accuracy: ~54.5%)
- Model 2 (Deep Neural Network): A complex architecture with 4 hidden layers. (Result: Failed due to Overfitting, ~50%)
- Model 3 (Regularised Network): Utilised high dropout rates to reduce noise. (Result: Underfitting, ~48%)
- Model 4 (LSTM/GRU Time-Series): A Recurrent Neural Network designed to capture multi-day momentum. (Accuracy: ~55.9%)

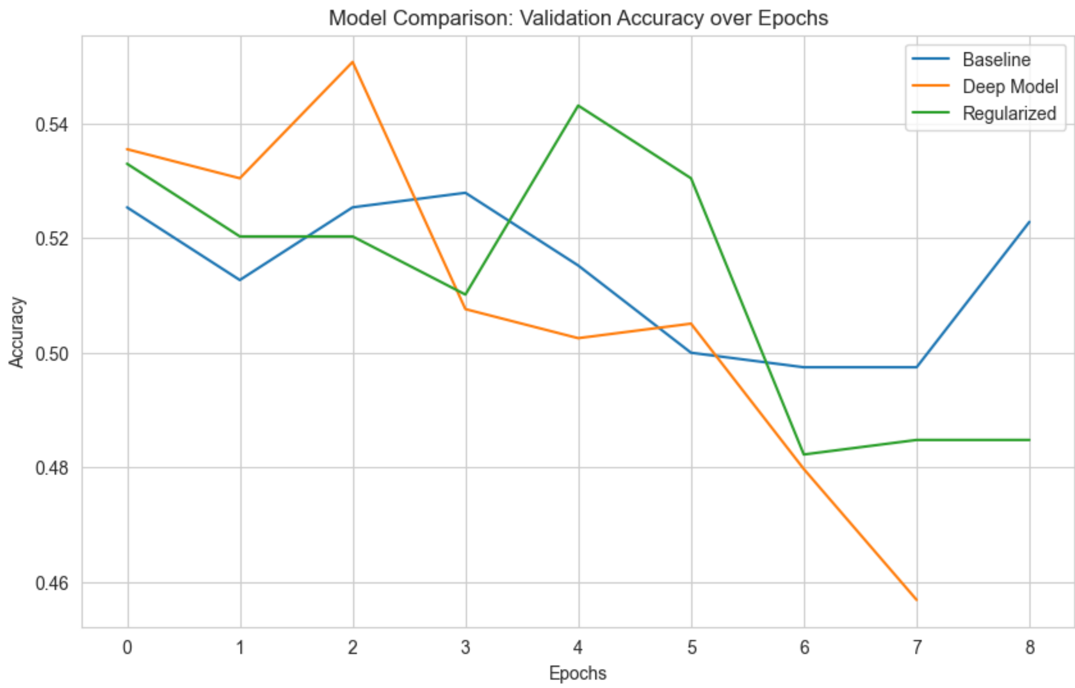

Figure 1: Standard deep learning models comparison

Selected Model: We selected an Ensemble Architecture (averaging the predictions of Model 1 and Model 4) as the final production model. Reasoning: This approach achieved the highest accuracy (57.52%) by combining the strength of "Spot Sentiment" analysis (Dense) with "Short-Term Momentum" (LSTM), successfully mitigating the variance found in individual models.

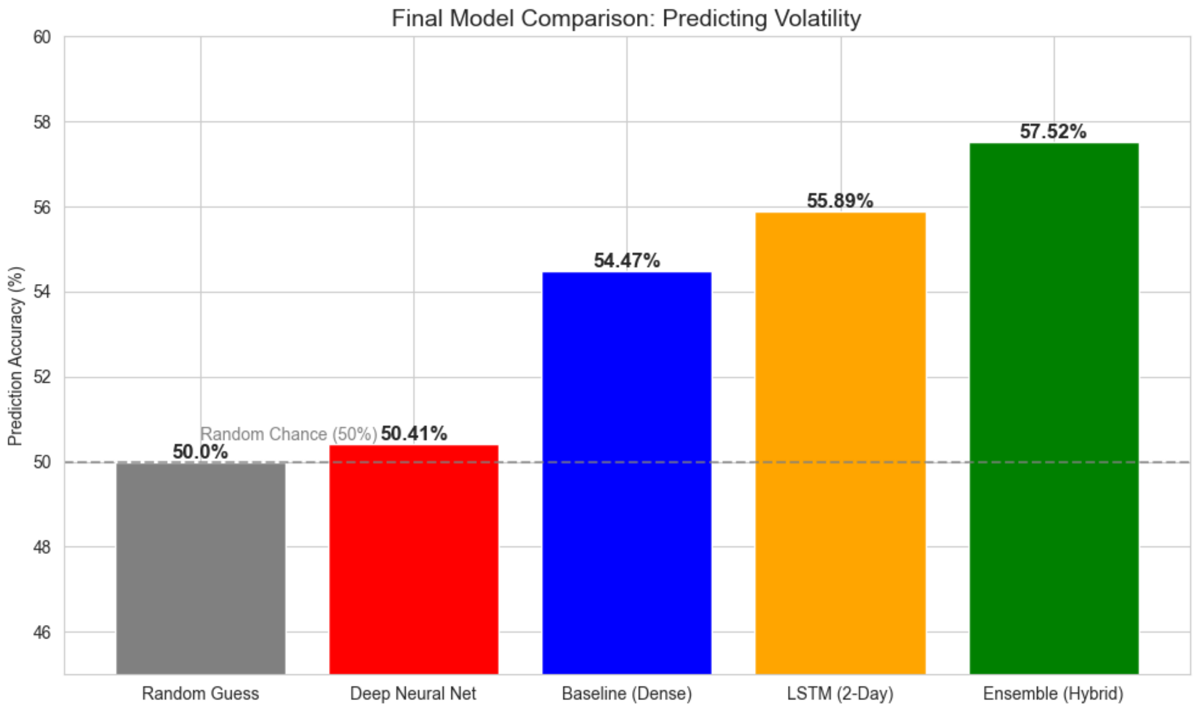

Figure 2: Final modles comparison

4. Key Findings

- **Content > Volume:** Simple tweet volume was a poor predictor. However, isolating specific topics (e.g., *Trade War* vs. *Election*) significantly increased predictive power, proving that the market reacts to the *context* of the message, not the frequency.
- **Market Efficiency:** The time-series analysis revealed a short "memory window" of approximately 48 hours. Tweets older than 2 days lost predictive value, confirming that social media sentiment is priced into the market almost instantly.
- **Measurable Edge:** The final model's accuracy of 57.52% represents a statistically significant edge over random chance (50%), validating the hypothesis that political NLP data contains actionable financial signal.

5. Limitations & Future Strategy

While successful, the current model highlighted specific limitations:

- Data Scarcity & Overfitting: Deep Learning typically requires massive datasets. With only ~1,000 trading days available, the deeper models (Model 2) suffered from rapid overfitting, memorising noise rather than learning patterns.
- Binary Simplification: Predicting simply "Up" or "Down" loses the nuance of *magnitude*. A tweet might cause a massive crash, but the model treats it the same as a small dip.

Plan of Action: To address these flaws, future iterations will:

1. Expand Data: Incorporate intraday pricing (hourly VIX) to multiply the number of training samples by 7x.
2. Granular Sentiment: Implement "Sentiment Polarity" (Positive/Negative scores) rather than just topic counts.
3. Regression Modelling: Shift from Binary Classification to Regression to predict the *magnitude* of the volatility spike, not just the direction.