2025 IEEE International Conference on eScience (eScience)

CryptexLLM: How LLM Generalizability Forecasts High Volatility

Logan Rao
Department of Computer Science
Illinois Institute of Technology
Chicago, USA
lrao1@hawk.illinoistech.edu

Mary Tsaryk
Department of Computer Science
Fordham University
New York City, USA
mt12@fordham.edu

Anwar Benhnini
Department of Computer Science
Illinois Institute of Technology
Chicago, USA
abenhnini@hawk.illinoistech.edu

Ioan Raicu
Department of Computer Science
Illinois Institute of Technology
Chicago, USA
iraicu@illinoistech.edu

ional Conference on eScience (eScience) | 979-8-3315-9145-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/eScience65000.2025.00071

***Abstract*—We present CryptexLLM, an approach to time series forecasting that adapts large language models (LLMs) for predicting high volatility data. We extend the TimeLLM framework with an adaptive weighted loss function, feature engineering, and sentiment analysis. Our experiments show that the approach we took outperforms traditional LSTM models and statistical methods, with our best performing model being Llama 3.1. The adaptations we made improve directional accuracy, which is particularly useful for financial applications, while still maintaining computational efficiency. Our results suggest that LLMs have the potential to effectively generalize to volatile time series domains.**

***Index Terms*—time series, volatile, llm**

I. Introduction

Time series forecasting is seen in many different domains, from electricity demand and web traffic monitoring to financial markets and healthcare diagnostics. While many time series datasets exhibit predictable seasonal patterns, high volatility data remains a challenge that traditional forecasting methods often fail at. Current state-of-the-art techniques tend to overfit to their specific training and struggle to generalize across different time series datasets. We address these limitations by adapting large language models (LLMs) for time series forecasting. We focus on Bitcoin price (BTC) prediction for

learned translation layer that consists of a patch embedder and multi-head attention mechanism. These components transform short numerical sequences into representations that a frozen LLM can process, whose output then gets re-transformed into a numerical prediction.

B. Experimental Setup

We evaluated six open-source LLMs: Llama 7B, Llama 3.1 8B, DeepSeek R1 7B, Mistral 7B, Gemma 3 1B, and Qwen 3 7B. Training utilized 4 nodes of 32GB V100 GPUs, with MLflow tracking experiments and MinIO object storing models. We used Optuna to optimize hyperparameters including input/output length, vocabulary size, patch size, learning rate, and forecasting mode (univariate/multivariate).

C. Feature Engineering

We augmented raw price data with technical indicators. These span momentum (RSI, SMA), volatility (Bollinger Bands, GARCH), and market behavior (VWAP, lag features). A correlation-based feature selection algorithm ranked features by their correlation to BTC price and removed redundant features to ensure computational time stays low while predictive efficiency stays high. Additionally, we integrated sentiment