Financial Forecasting Using Deep Learning and Text Data: A Cryptocurrency Return Prediction Case Study

Duygu Ider, Vincent Gurgul, Stefan Lessmann[1]

Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

Anticipating price developments in financial markets is a topic of continued interest in forecasting. Funneled by advancements in deep learning and natural language processing (NLP) together with the availability of vast amounts of textual data in form of news articles, social media postings, etc., an increasing number of studies incorporate text-based predictors in forecasting models. The hypothesis that text-based cues carry predictive information appears especially plausible for cryptocurrency markets, which are the subject of this study. Noting some variation in how specifically prior work employs text features and advanced pre-trained NLP models for price modeling, the first research goal is to survey available options and systematically compare selected approaches. For example, many studies convert raw text into sentiment signals and use these as auxiliary inputs in a financial forecasting model. Concentrating on sentiment and, thereby, disregarding the actual content of the text, one may ask whether the sentiment extraction approach is suitable. In this context, we emphasize weak labeling, an NLP approach to finetune text classifiers on an unlabeled target corpus, which holds potential for financial forecasting but has received little attention in the corresponding literature. A second research goal concerns the well-known curse of dimensionality. Combining fundamental, technical, and text-based features in a forecasting model increases dimensionality substantially. This challenge is amplified in cryptocurrency forecasting in that yet another set of potentially relevant features can be extracted from the blockchain. We examine the predictive value of the different feature groups and benchmark alternative forecasting methods for high-dimensional time series data. In this context, we emphasize the temporal fusion transformer, which promises automatic feature selection, and assess its effectiveness vis-a-vis benchmark forecasting models. In sum, our study contributes original empirical evidence to the literature on cryptocurrency price forecasting. We also introduce several advanced deep learning methodologies for text processing and time series forecasting to the financial market modeling community, which are generally applicable in the field and can benefit financial forecasting at large.

---

[1] Corresponding author, E-Mail: stefan.lessmann@hu-berlin.de, Tel. +40.30.2093.99540