NED University of Engineering & Technology

Masters Thesis Information Showcase
Spring Semester-2026
Year 2026

	Department: Mathematics Programme: Choose an item. Specialization: Choose an item.
1	**Title of the Thesis** Integrating sentiment analysis into financial market forecasting
2	**Abstract** (min 300 to 500 words) Cryptocurrency markets are highly volatile, nonlinear, and strongly influenced by historical price behavior, technical patterns, news events, and investor sentiment. This study investigates the integration of sentiment analysis into financial market forecasting by developing a comprehensive machine and deep learning framework for predicting Bitcoin price direction and magnitude. Financial news headlines and social media tweets are utilized to capture market sentiment, while historical Bitcoin prices and technical indicators provide quantitative information regarding market behavior. Sentiment is extracted using VADER, a lexicon-based approach, and FinBERT, a domain-specific transformer model designed for financial text. The extracted sentiment features are combined with technical indicators, including lagged prices, moving averages, percentage change, and Relative Strength Index. Multiple machine learning models, including Logistic Regression, Random Forest, XGBoost, CatBoost, and Gradient Boosting, are comparatively evaluated alongside deep learning architectures comprising CNN, LSTM, GRU, GRU-CNN, Meta-SD, and Meta-SGD. Both classification and regression experiments are conducted to predict market direction and continuous cryptocurrency price levels, respectively. Model performance is assessed under random and chronological time-series splitting strategies to examine predictive capability under conventional and realistic forecasting conditions. The findings demonstrate that sentiment contains limited but measurable information for directional forecasting, with several classification models achieving performance moderately above random guessing. However, sentiment-only features demonstrate weak explanatory capability for continuous price prediction, particularly under time-series validation. Incorporating technical indicators substantially improves regression performance, indicating that historical price behavior, momentum, and market trends provide stronger predictive information than sentiment alone. Results further reveal that random splitting can produce overly optimistic estimates, whereas chronological validation exposes non-stationarity and changing market regimes. Cross-cryptocurrency experiments demonstrate broadly consistent patterns, supporting framework generalizability. Overall, reliable cryptocurrency forecasting requires the combined consideration of sentiment information, technical indicators, appropriate predictive models, and time-aware validation. The framework provides useful methodological insights for researchers, investors, and financial analysts developing intelligent, robust, and realistic data-driven forecasting systems.
3	**Impact on SDG#___"** (min 500 words) (Please mention the impact of the relevant attained SDG)

SDG 9: Industry, Innovation and Infrastructure

The proposed research contributes to **United Nations Sustainable Development Goal 9 (SDG 9): Industry, Innovation and Infrastructure** by advancing the application of artificial intelligence, machine learning, deep learning, and natural language processing to intelligent financial analytics. Modern financial markets increasingly depend on digital infrastructure, computational technologies, and data-driven decision-support systems. The rapid development of cryptocurrency markets has further increased the need for reliable analytical frameworks capable of processing heterogeneous information and supporting informed financial decisions.

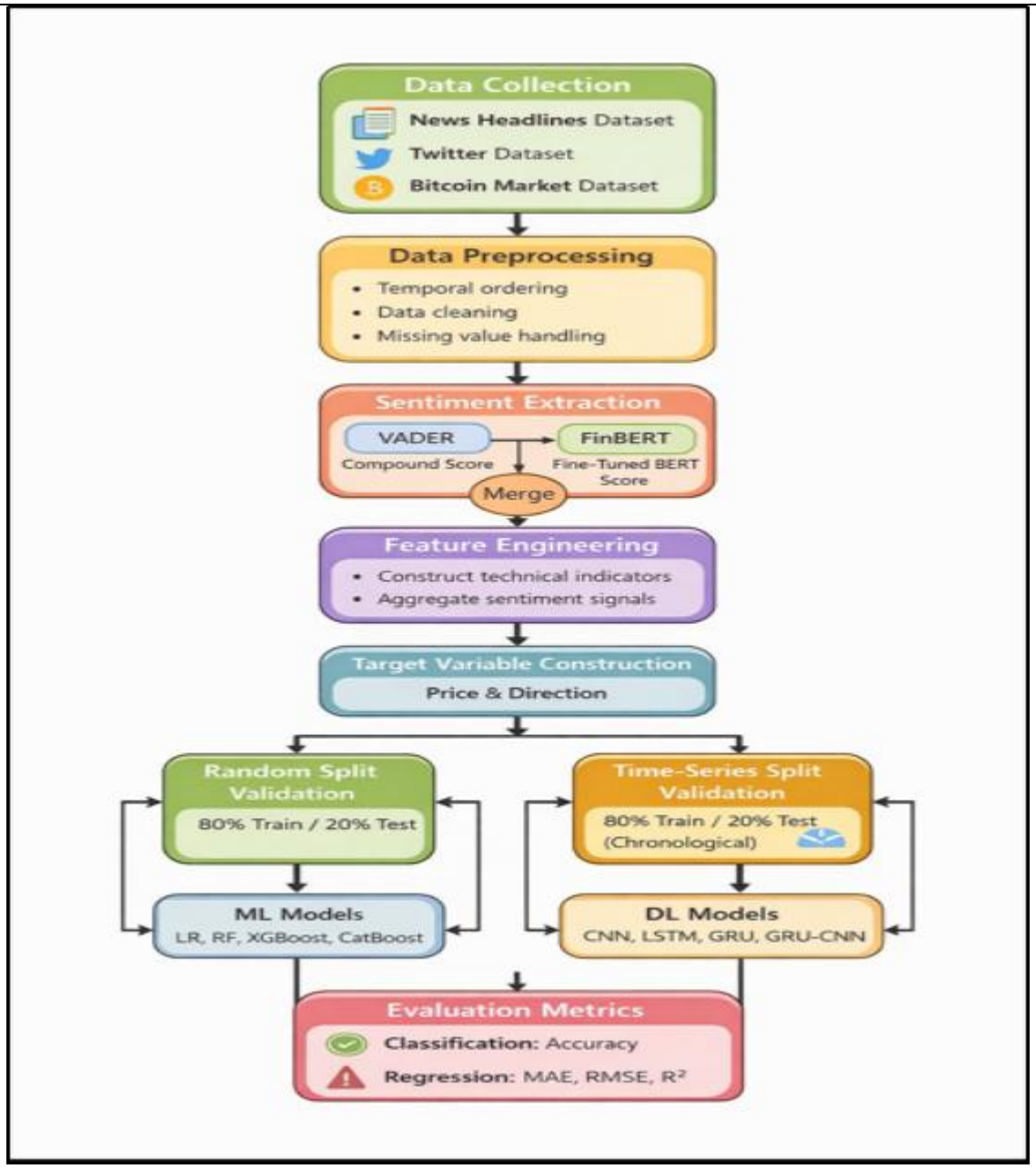
The study contributes to technological innovation by integrating **structured and unstructured financial information within a unified predictive framework**. Rather than relying exclusively on historical cryptocurrency prices, the research incorporates financial news headlines, social media tweets, historical Bitcoin data, and technical market indicators. This integration enables the forecasting framework to examine both quantitative market behavior and qualitative investor sentiment. The thesis specifically identifies this multi-source integration as an important contribution to cryptocurrency forecasting.

A further innovative component is the use of two different sentiment-analysis approaches: **VADER and FinBERT**. VADER provides lexicon-based sentiment extraction, while FinBERT represents a transformer-based language model specialized for financial text. Their integration with machine learning and deep learning models demonstrates how contemporary NLP technologies can be incorporated into financial analytics to extract potentially useful information from large volumes of textual data.

The study also supports SDG 9 through its comparative application of advanced computational models. Machine learning approaches such as Logistic Regression, Random Forest, XGBoost, CatBoost, and Gradient Boosting are evaluated alongside CNN, LSTM, GRU, GRU-CNN, Meta-SD, and Meta-SGD deep-learning architectures. This comparative framework contributes to research capacity by examining how different computational architectures respond to sentiment information, technical indicators, and changing market conditions.

Importantly, the research does not focus solely on maximizing reported predictive accuracy. It investigates the **reliability of AI-based financial forecasting** by comparing random data splitting with chronological time-series validation. The findings demonstrate that random splitting may generate overly optimistic performance because it does not adequately represent temporal forecasting conditions. Time-series validation provides a more realistic assessment by preserving chronological order. This methodological contribution is particularly important for responsible financial AI because apparently high model accuracy has limited practical value if it does not generalize to future market conditions.

	The research further demonstrates that sentiment information alone is insufficient for reliable cryptocurrency price-level prediction. Instead, technical indicators—including lag variables, moving averages, percentage changes, and RSI—provide more stable predictive information. Cross-cryptocurrency validation also produces patterns broadly consistent with the primary Bitcoin experiments, strengthening the methodological relevance of the framework beyond a single digital asset. Overall, this research supports **SDG 9** by advancing innovative, AI-driven infrastructure for financial data analysis and forecasting. Its impact lies not merely in developing predictive models but in promoting **more rigorous, realistic, and reliable financial AI systems** through multi-source data integration, NLP-based sentiment analysis, technical feature engineering, machine and deep learning, and time-aware validation. Such methodologies can contribute to stronger digital financial analytics and provide useful decision-support insights for researchers, investors, financial analysts, and institutions operating within increasingly data-driven financial environments.
4	**Scholar Name** (along with contact details) Sana Aslam aslamsana771@gmail.com,
5	**Supervisor & Co-supervisor Name** (along with contact details) Dr. Komal Batool Dr. Ubaida Fatima
6	**Graphical Abstract**


Data Collection
News Headlines Dataset
Twitter Dataset
Bitcoin Market Dataset
Data Preprocessing
Temporal ordering
Data cleaning
Missing value handling
Sentiment Extraction
VADER
Compound Score
FinBERT
Fine-Tuned BERT Score
Merge
Feature Engineering
Construct technical indicators
Aggregate sentiment signals
Target Variable Construction
Price & Direction
Random Split Validation
80% Train / 20% Test
Time-Series Split Validation
80% Train / 20% Test (Chronological)
ML Models
LR, RF, XGBoost, CatBoost
DL Models
CNN, LSTM, GRU, GRU-CNN
Evaluation Metrics
Classification: Accuracy
Regression: MAE, RMSE, R²