Faculty of Technology
University of Sri Jayewardenepura

ITS 4202
Emerging Technologies
Assignment 3

Group Members

G.S. Chamika - ICT/20/818

Y. N. S. Dissanayake - ICT/20/837

B. G. D. T. T. Jeerasinghe - ICT/20/863

S.A. Dilanka Sandeepa - ICT/20/926

Contents

1. Steps Followed for Prompt Engineering

1.1 Journal Papers

Paper 1: Bitcoin price prediction using machine learning: An approach to sample dimension engineering
https://doi.org/10.1016/j.cam.2019.112395

Paper 2: Prediction of Bitcoin Price Using Bi-LSTM Network
https://doi.org/10.1109/ICCCI50826.2021.9402427

Paper 3: Bitcoin Price Prediction Using Machine Learning and Deep Learning Algorithm
https://doi.org/10.1109/ICRITO56286.2022.9964677

1.2 Understanding the Task Requirements

1. Identify research gaps across three selected papers.

2. Compare methodologies, key findings, limitations, and contradictions.

3. Use two AI tools separately: ChatGPT and DeepSeek.

4. Select relevant sections from each paper to ensure a focused analysis.

1.3 Selecting Relevant Subsections

To ensure that AI analyzes meaningful content, we selected the following key sections from each journal paper.

- **Abstract**: Provides a high-level summary of each paper (objectives, methodology, and key findings).
- **Literature Review**: Helps understand what prior work has been done.
- **Results and Discussion**: Contains the main experimental findings, model performance, and key conclusions.
- **Conclusion**: Summarizes key findings and suggests future research.

1.4 Designing an Effective AI Prompt

To ensure high-quality responses from AI tools, we structured the prompt using best practices in prompt engineering. Below is the detailed approach.

A. Clearly Define the Role of AI

To make AI responses accurate and insightful, we instructed the AI to assume a specific role.

Prompt: "Assume you are an AI research assistant specializing in financial technology and cryptocurrency price prediction models."

B. Provide Context and Background

To help AI understand the scope of the task, we included relevant background details.

Prompt: "You have been provided with three journal papers on Bitcoin price prediction using machine learning and deep learning techniques. These papers focus on statistical models, deep learning architectures, and hybrid approaches for forecasting Bitcoin prices."

Contextual Details Included

- Machine Learning & Deep Learning Approaches (e.g., ARIMA, LSTM, XGBoost).
- Different evaluation metrics (MAE, RMSE, Accuracy, MAPE).
- Data sources and periods for model training.
- Challenges in cryptocurrency prediction, such as high volatility and market dynamics.

C. Ask for Specific Outputs

Prompt: "Analyze the provided journal papers and provide a structured comparison based on the following key aspects."

1. Methodologies Used
2. Key Findings & Accuracy
3. Limitations and Inconsistencies
4. Research Gaps
5. Future Research Directions

D. Encourage Critical Thinking

To push AI towards a deeper analysis, we added instructions to think like an expert.

Prompt: "Evaluate contradictions in findings and suggest how future research can resolve these issues."

Examples of contradictions AI should analyze

- Statistical vs. Deep Learning Models

ARIMA performs well for time-series data, but LSTM is better for capturing volatility.

- Evaluation of Metric Discrepancies

Some studies use accuracy, while others use MAE/MAPE, making direct comparisons difficult.

1.5 Final Prompt Used

"Assume you are an AI research assistant specializing in financial technology and cryptocurrency price prediction models.

You have been provided with three journal papers on Bitcoin price prediction using machine learning and deep learning techniques. These papers focus on statistical models, deep learning architectures, and hybrid approaches for forecasting Bitcoin prices.

Task Instructions

Analyze the provided journal papers and provide a structured comparison based on the following key aspects:

1. Compare Methodologies
    - Compare statistical models (ARIMA, Regression) with ML/DL approaches (LSTM, Bi-LSTM, XGBoost).
    - Discuss data preprocessing techniques and feature selection.

2. Analyze Key Findings & Accuracy
    - Compare model performance across different datasets.
    - Identify strengths and weaknesses of each approach

3. Identify Limitations and Inconsistencies

- Highlight methodological contradictions.
- Discuss dataset limitations, feature selection issues, and evaluation metric discrepancies.

4. Extract Research Gaps

- Identify missing elements and unexplored areas.
- Discuss the lack of real-time adaptability and hybrid modeling approaches.

5. Propose Future Research Directions

- Suggest potential improvements, such as hybrid models (ARIMA-LSTM), reinforcement learning, and macro-economic feature integration.

Output Format

- Provide the response in a **structured bullet-point format**.
- Use **comparative tables where relevant**.
- Highlight contradictions in findings.
- Suggest how future research can resolve existing limitations.

Critical Thinking Requirement

- Evaluate contradictions in findings.
- Justify why certain models work better than others.
- Recommend how future studies can standardize feature selection and evaluation metrics.

By following these instructions, ensure that the response is analytical, structured, and focused on identifying research gaps."

2. Output Obtained Through ChatGPT

1. Comparison of Methodologies

Aspect	Paper 1	Paper 2	Paper 3
Models Used	Logistic Regression, Linear Discriminant Analysis, Random Forest, XGBoost, Quadratic Discriminant Analysis, Support Vector Machine (SVM), and LSTM.	Bidirectional Long Short-Term Memory (Bi-LSTM) for sequential forecasting.	ARIMA, LSTM, GRU, Ridge Regression, LASSO Regression, Linear Regression.
Dataset	Daily and 5-minute Bitcoin price data from CoinMarketCap (2017-2019).	Bitcoin historical price data from Kaggle (2012-2020).	Bitcoin market data from CryptoCompare API (4-month dataset).
Feature Engineering	High-dimensional features including investor sentiment, Google Trends, gold spot price, and trading volume.	Time-series forecasting approach with minute-level data aggregated into daily format.	Focuses mainly on price indicators like closing price, high/low values, and volume.
Evaluation Metrics	Accuracy (66% for daily data using statistical models, 67.2% for high-frequency data using ML models).	Mean Absolute Percentage Error (MAPE = 13%).	Mean Absolute Error (MAE) comparison, with ARIMA achieving the best results (MAE: 0.02294).

2. Key Findings & Accuracy Analysis

Paper 1

- Logistic Regression achieved 66% accuracy for daily price predictions.
- LSTM models were more effective for high-frequency (5-minute interval) price predictions, reaching 67.2% accuracy.
- Feature engineering plays a crucial role in model performance.

Paper 2

- Bi-LSTM model captured sequential trends in Bitcoin prices.
- Achieved a MAPE of 13%, outperforming traditional regression-based models.
- Demonstrated that bidirectional processing enhances predictive accuracy compared to unidirectional LSTM.

Paper 3

- ARIMA model performed best in time-series forecasting, with the lowest MAE (0.02294).
- Other ML and DL models (LSTM, GRU) had higher errors, indicating that ARIMA is more effective for structured time-series analysis.

3. Limitations & Inconsistencies

Limitation	Paper 1	Paper 2	Paper 3
Feature Selection	Uses a wide range of features but does not include real-time news sentiment or macroeconomic indicators.	Does not incorporate external factors such as social sentiment or economic data.	Rely on market-based price indicators, limiting interpretability.
Data Scope	The dataset ends in 2019, excluding post-2020 market volatility (e.g., COVID-19 effects).	Minute-level data was converted to daily, leading to a loss of intraday trading patterns.	Uses only a 4-month dataset, reducing long-term predictive performance.
Methodological Contradictions	Statistical models outperformed ML for daily price prediction, but ML worked better for high-frequency data.	Bi-LSTM performed well but lacked comparison with statistical models (e.g., ARIMA).	ARIMA outperformed ML/DL models, contradicting claims that deep learning is superior.

4. Research Gaps Identified

1. Lack of Hybrid Approaches

- Paper 1 focuses on statistical and ML models but does not combine them into a hybrid approach.
- Paper 2 only evaluates Bi-LSTM, without comparing it with statistical models.
- Paper 3 compares ML/DL models but lacks an ensemble model for combining ARIMA and LSTM.

2. Limited Use of External Factors

- Only Paper 1 considers investor sentiment and Google Trends.
- No study incorporates macroeconomic indicators, regulatory policies, or social media sentiment analysis.
- Future models should integrate Twitter sentiment, news data, and global economic trends.

3. Handling Market Volatility

- Paper 2's Bi-LSTM struggles with extreme Bitcoin price swings.
- Paper 3's ARIMA assumes a stationary time series, limiting its ability to adapt to volatile trends.
- Future research should use GARCH, Bayesian inference, or volatility-adjusted deep learning models.

4. Real-Time Prediction & Adaptive Learning

- None of the studies implement real-time forecasting models.
- Future approaches should integrate reinforcement learning and transformers for real-time adaptability.

5. Proposed Future Research Directions

1. Hybrid ML-DL Model for Bitcoin Forecasting

- Combine ARIMA for trend analysis with LSTM/Bi-LSTM for short-term price fluctuations.
- Use ensemble learning techniques (e.g., Stacking, Bagging, Boosting).

2. Incorporation of Market Sentiment & External Factors

- Leverage Twitter sentiment, financial news, Google Trends, and blockchain activity metrics.
- Integrate macroeconomic indicators like inflation rates, interest rates, and stock market trends.

3. Real-Time & High-Frequency Adaptability

- Develop real-time adaptable models using reinforcement learning and transformers.
- Incorporate live streaming data for more accurate price predictions.

4. Standardization of Feature Selection & Metrics

- Establish a benchmark dataset with uniform evaluation metrics (RMSE, MAPE, MAE, Sharpe Ratio).
- Standardize data handling between high-frequency and daily datasets.

5. Addressing Market Volatility

- Implement volatility clustering models (GARCH, ARCH) with LSTM models.
- Explore probabilistic forecasting techniques to capture uncertainty.

3. Output Obtained Through DeepSeek

1. Methodologies Comparison

Aspect	Paper 1	Paper 2	Paper 3
Models	- Statistical: LR, LDA - ML/DL: RF, XGBoost, QDA, SVM, LSTM	DL: Bi-LSTM	- Statistical: ARIMA - ML: Linear/Ridge/LASSO - DL: LSTM, GRU
Data Preprocessing	- Feature engineering (high-dimensional for daily, low-dimension for 5-min data)	- Windowing (50-day window) - Normalization	- Time-series differencing (ARIMA) - Train-test split (80-20)
Feature Selection	- Integrated gold price, Google Trends, network metrics	- Basic trading data (open/close/high/low/volume)	- Limited to OHLCV (open/high/low/close/volume)
Data Granularity	- Daily vs. 5-minute intervals	- Daily data	- Daily data

2. Key Findings & Accuracy

Model	Paper 1	Paper 2	Paper 3
Best Accuracy	- Daily: LR (66%) - 5-min: LSTM (67.2%)	- Bi-LSTM (MAPE: 13%)	- ARIMA (MAE: 0.02294)
Strengths	- Tailored models for data granularity - High-dimensional features improve daily predictions	- Captures bidirectional temporal dependencies	- ARIMA excels in time-series data
Weaknesses	- Accuracy metric unsuitable for regression tasks - Limited DL exploration for daily data	- No comparison with statistical models - Small dataset (2017–2019)	- Small dataset (4 months) - Ignores external features

Contradictions:

- ARIMA vs. LSTM: Paper 3 found ARIMA superior (MAE: 0.02294 vs. LSTM: 0.02821), while Paper 1 showed LSTM outperformed statistical methods for high-frequency data (67.2% accuracy).
- Feature Impact: Paper 1 highlights external features (e.g., gold price) as critical, while Papers 2–3 rely solely on trading data.

3. Limitations & Inconsistencies

Issue	**Details**
Methodological Gaps	- Paper 1 uses accuracy for regression (misleading metric). - Paper 3 excludes external features (e.g., sentiment).
Dataset Limitations	- Paper 3: 4-month dataset → risk of overfitting. - Paper 2: No high-frequency data.
Evaluation Metrics	- Inconsistent metrics: Accuracy (Paper 1) vs. MAE/MAPE (Papers 2–3).
Feature Selection Bias	- Paper 1 uses Boruta for feature selection; others use manual selection.

4. Research Gaps

- Hybrid Models: No paper combines statistical (e.g., ARIMA) and DL models (e.g., LSTM) for joint linear-nonlinear pattern extraction.
- Real-Time Adaptability: All models use historical data; none test real-time streaming or adaptive learning.
- Macro-Economic Features: Only Paper 1 integrates gold price and Google Trends; others ignore macroeconomic/sentiment factors.
- Standardization: No consensus on evaluation metrics (accuracy vs. MAE vs. MAPE) or dataset sizes.

5. Future Research Directions

Area	**Recommendations**
Hybrid Models	- Combine ARIMA (linear trends) with LSTM (nonlinear patterns) for improved robustness.
Feature Expansion	- Integrate macroeconomic indicators (e.g., inflation rates), social media sentiment, and blockchain metrics (e.g., hash rate).
Real-Time Adaptation	- Test reinforcement learning (RL) for dynamic model updates in volatile markets.
Standardization	- Adopt uniform metrics (e.g., RMSE, MAE) and benchmark datasets for comparability.
Dataset Diversity	- Use larger, multi-year datasets spanning bull/bear markets to avoid temporal bias.

Critical Analysis & Recommendations

- Why ARIMA Works: Paper 3's ARIMA excels due to its focus on stationary time-series data, while LSTM/GRU underperforms without sufficient training data.
- Contradiction Resolution: Standardized datasets and metrics (e.g., MAE for regression) would clarify model superiority across contexts.
- Feature Engineering: Future studies should automate feature selection (e.g., using SHAP values) to reduce bias and improve reproducibility.

4. Comparative Report: Identifying Research Gaps in Bitcoin Price Prediction Studies

2.1 Introduction

This report compares the research gap identification results obtained from two AI-assisted tools: ChatGPT and DeepSeek. The goal is to evaluate their performance in analyzing methodologies, key findings, limitations, inconsistencies, and future research directions in Bitcoin price prediction studies.

Objectives of the Report

- Compare the methodologies, key findings, and limitations identified by both tools.
- Analyze the research gaps extracted by each tool.
- Evaluate the strengths and weaknesses of each AI-generated response.
- Discuss how future research can address identified gaps.

2.2 Structural Comparison of AI Responses

Aspect	ChatGPT	DeepSeek
Organization	Structured with clear headings and subsections for methodologies, key findings, limitations, research gaps, and future directions.	Structured but more condensed, making it slightly harder to differentiate research gaps from limitations.
Clarity & Readability	Uses tables, bullet points, and summaries to enhance readability.	More compact text, making it harder to navigate quickly.
Comparative Analysis Depth	Provides detailed comparisons of models, datasets, and evaluation metrics.	Efficient summary but lacks detailed discussion on contradictions and methodology variations.

Observations

- ChatGPT's structured breakdown makes it easier to compare findings and limitations across the three papers.
- DeepSeek provides a denser summary, making it more concise but less detailed.

2.3 Methodological Comparison

Criterion	ChatGPT	DeepSeek
Models Compared	Covers ARIMA, LSTM, Bi-LSTM, XGBoost, and Regression models, and evaluates their effectiveness in different scenarios.	Discusses similar models but focuses heavily on ARIMA vs. LSTM, giving less attention to other ML models.
Feature Engineering Consideration	Discusses external factors (Google Trends, gold price, social sentiment, macroeconomic indicators) in-depth.	Mentions external factors briefly but does not elaborate on macroeconomic influences.
Data Granularity Analysis	Highlights the importance of minute, hourly, and daily granularity in forecasting accuracy.	Acknowledges data granularity but does not explore solutions for handling different timeframes.

Observations

- ChatGPT provides a broader discussion, ensuring different models are well-compared.
- DeepSeek focuses more on ARIMA vs. LSTM, missing hybrid approaches, and ML feature engineering discussions.

2.4 Identification of Research Gaps

Research Gap	ChatGPT	DeepSeek
Hybrid Models	It is recommended that ARIMA be combined with LSTM/Bi-LSTM for better predictive accuracy.	Mentions hybrid models but lacks details on ensemble learning approaches.
Real-Time Adaptability	Emphasizes the need for reinforcement learning and transformers for real-time data integration.	Acknowledges real-time challenges but does not suggest specific models.
Feature Selection & Expansion	Proposes Twitter sentiment, financial news, and blockchain activity metrics for improved forecasting.	Briefly mentions feature selection but does not discuss blockchain data integration.
Standardization of Evaluation Metrics	Recommend a benchmark dataset with RMSE, MAPE, MAE, and Sharpe Ratio.	Notes inconsistencies in metrics but does not propose a standardized evaluation framework.

Observations

- ChatGPT covers a wider range of research gaps and suggests detailed solutions.
- DeepSeek identifies key research gaps but does not propose in-depth resolutions.

2.5 Contradiction Resolution

Contradiction	ChatGPT	DeepSeek
ARIMA vs. LSTM Superiority	Explains why ARIMA is better for structured time series, while LSTM is better for volatile trends.	Mentions that ARIMA performed better but does not discuss the context-dependent effectiveness.
Feature Selection Differences	Discusses why Papers 2 & 3 ignore external factors while Paper 1 includes them.	Mentions that Papers 2 & 3 lack external features but do not provide analysis on why.

Evaluation Metrics	Highlights inconsistent evaluation of metrics across papers and recommends standardized benchmarks.	Mentions inconsistencies but does not propose a universal standard.

Observations

- ChatGPT provides a stronger critical analysis, explaining why contradictions exist and how to resolve them.
- DeepSeek acknowledges contradictions but does not explore resolutions in depth.

2.6 Future Research Directions

Future Direction	ChatGPT	DeepSeek
Hybrid Models	Suggests stacking ARIMA and LSTM models for improved forecasting.	Recommends hybrid approaches but without specific implementation suggestions.
Feature Expansion	Proposes macroeconomic integration (inflation, stock indices, social sentiment, blockchain metrics).	Mentions macroeconomic indicators but lacks detail on how they can be integrated.
Real-Time Adaptability	Recommends reinforcement learning and real-time transformers for live prediction models.	Acknowledges real-time issues but does not propose solutions.
Evaluation Standardization	Advocates for a benchmark dataset with standardized metrics (MAE, RMSE, MAPE, Sharpe Ratio).	Notes metric inconsistencies but does not suggest standardization.

Observations

- ChatGPT proposes specific, actionable recommendations.
- DeepSeek highlights general research directions but lacks implementation details.

2.7 Conclusion

Both ChatGPT and DeepSeek successfully identified research gaps, but ChatGPT provided deeper critical analysis, contradiction resolution, and future research recommendations. DeepSeek is more concise but lacks depth in methodology comparisons and solutions.

Future research should focus on

- Hybrid models (ARIMA + LSTM)
- Integration of real-time adaptability
- Expanding feature selection with macroeconomic & blockchain data
- Standardizing evaluation metrics and datasets