UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Forecasting DeFi Markets: A Network-Based Approach to AAVE Price Dynamics

Author: Théo FEDER

Supervisor: Natkamon TOVANICH

Date: October 28, 2024

Abstract

This study integrates network-based approaches with machine learning techniques to forecast WETH and WBTC token price. Utilizing data from The Graph Platform, we analyze 878,601 daily market entries and 1,514,611 position snapshots from November 2020 to May 2023. Our methodology combines PageRank and centrality measures with LSTM networks. We identify key network structures influencing price dynamics, revealing patterns in user activity, interest rates, and risk management. The results demonstrate the significance of network effects in DeFi price prediction. This research contributes to understanding the interplay between network dynamics and price movements in DeFi ecosystems, offering insights for more informed decision-making in this rapidly evolving financial landscape.

Contents

1 Introduction

Decentralized finance (abbreviated as DeFi) refers to an alternative financial system built on a permissionless blockchain that promises openness, efficiency, transparency, interoperability, and decentralization [1, 2], where users can anonymously interact in a digital environment while preserving the benefits of cash such as privacy. AAVE is a decentralized bank that allows users to lend and borrow crypto assets and earn interest on assets supplied to the protocol [3]. Having experienced significant growth in recent years, the market value of AAVE was beyond 2.9 billion U.S. dollars as of Jan. 20, 2022 [4]. In general, decentralized banks like AAVE differ from centralized banks in two aspects: 1) they replace centralized credit assessments with coded collateral evaluation [5], and 2) they employ smart contracts to execute asset management automatically [6].

While it is widely acknowledged that traditional financial systems are vulnerable to contagion through various channels, including bank runs [7] and default cascades [8], little is known about the contagion risks potentially present in DeFi protocols. [9] studied financial contagion in Compound V2, a decentralized lending protocol, using stress tests to identify pools most likely to set off a cascade of defaults.

Complementing such risk analysis, this study aims to predict the token prices using network modeling techniques, enhancing our understanding of decentralized lending dynamics. By incorporating network effects into price forecasting, we seek to capture DeFi ecosystem interactions that traditional methods might overlook [10]. Our approach offer insights for optimizing liquidity provision and borrowing strategies, potentially contributing to liquidity risk mitigation [11]. The network-based analysis could also provide a new perspective on systemic risk assessment in DeFi [12].

Our research explores how AAVE token price predictions and network dynamics might inform protocol design, potentially contributing to DeFi system resilience [13]. We aim to generate data that could assist users and investors in decision-making, considering both price movements and network effects [14]. While ambitious, we acknowledge that improved price forecasting may only modestly impact market efficiency and volatility in the complex DeFi ecosystem [15].

2 Background

Unlike conventional financial institutions, which rely on central authorities to manage transactions and assess creditworthiness, DeFi leverages blockchain technology and smart contracts to facilitate peer-to-peer financial interactions without intermediaries [31]. This innovation enables users to engage in various financial activities, including lending, borrowing, trading, and earning interest on digital assets, in a trustless environment.

As of mid-2024, the market capitalization of DeFi lending protocols stands at approximately \$9.13 billion [4]. These protocols, which include AAVE, allow users to lend and borrow cryptocurrencies through smart contracts that automate the entire process [16]. These smart contracts ensure that transactions are executed according to predefined rules without the need for human intervention, thereby reducing the potential for human error and increasing efficiency.

Smart contracts, self-executing contracts with terms encoded directly into the blockchain, manage tasks such as collateralization, interest calculation, and automatic liquidation. This automation reduces human error and ensures precise execution of transactions [28]. Platforms like AAVE require over-collateralization to mitigate default risks, where borrowers provide collateral exceeding the loan value, ensuring continuous monitoring and automatic liquidation if collateral values fall below a threshold [6].

Interest rates in decentralized lending are determined by supply and demand dynamics, reflecting current market conditions and providing fair rates for lenders and borrowers [30]. The transparency of decentralized platforms, with all transactions recorded on a public ledger, builds user trust and enables independent verification of transactions [18]. These platforms are accessible to anyone with an internet connection and a digital wallet, democratizing financial services, particularly in regions with underdeveloped banking infrastructure [19]. The elimination of intermediaries also offers faster transactions and lower fees, enhancing overall efficiency [5].

However, decentralized lending is not without risks. Smart contracts, while efficient, are susceptible to bugs and exploits, potentially leading to significant financial losses [16]. The evolving regulatory environment for DeFi introduces uncertainty that could impact the growth and stability of these platforms [17]. Furthermore, the value of collateral, often tied to volatile cryptocurrencies, can trigger liquidations and market instability during sharp price declines [2].

3 Related work

3.1 Network dynamics in financial markets

Although limited, studies have addressed the systemic risks and operational mechanisms of DeFi, particularly in lending protocols like AAVE.

[20] provide insights into the economic mechanisms and systemic risks in DeFi lending protocols, highlighting the importance of collateralization and its implications for AAVE's stability . [2] offers an extensive overview of DeFi, emphasizing the role of smart contracts in ensuring transparency and efficiency, which is crucial for AAVE's operations. [21] focuses on leverage and liquidity risks in DeFi platforms, emphasizing the necessity of robust risk management frameworks, directly applicable to AAVE's reliance on collateral and liquidity.

[9] investigate contagion risks in DeFi using stress tests on Compound V2, revealing insights into potential default cascades. This serves as an example of how network modeling can be effectively applied in this context, as shown below.

Particularly, network analysis has been used to study the impact of liquidity provision and borrowing activities on DeFi token price stability, suggesting network centrality as a reliable indicator of price volatility and systemic risks [32].

3.2 Price forecasting models in DeFi

Price forecasting in decentralized finance (DeFi) is essential for optimizing strategies, mitigating risks, and enhancing efficiency. Various advanced methodologies are utilized to predict price dynamics, including machine learning (ML), network-based models, hybrid models, and sentiment analysis.

Machine learning models, such as time series analysis, regression models, and neural networks, process vast amounts of historical data to identify complex patterns. For example, deep learning techniques have shown significant accuracy in predicting cryptocurrency prices [30]. LSTM networks combined with attention mechanisms have been used to forecast Bitcoin prices, capturing market trends and fluctuations effectively [31].

Hybrid models combine machine learning and network analysis to enhance predictive power. For instance, network features integrated into a gradient boosting machine framework have outperformed standalone ML models in predicting DeFi token prices [33].

4 Data and descriptive statistics

AAVE is a decentralized finance (DeFi) protocol that allows users to lend and borrow cryptocurrencies without intermediaries. Launched in January 2020, AAVE has quickly grown to become one of the largest and most popular DeFi platforms. As of January 2022, the market value of AAVE was beyond 2.9 billion U.S. dollars, reflecting its significant adoption and use in the DeFi ecosystem.

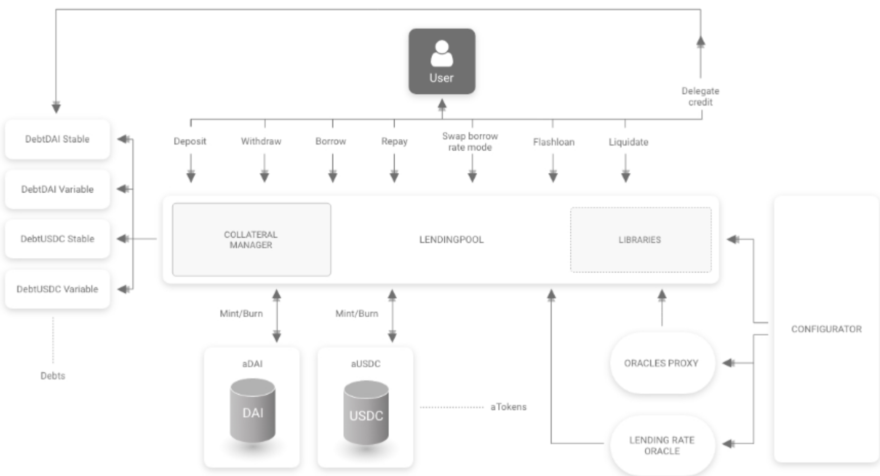

Figure 1: The diagram illustrates the components and interactions within the Aave V2 protocol. Users can deposit, withdraw, borrow, repay, and perform other actions via the Collateral Manager and Lending Pool. Key elements include aTokens (representing deposits), Debt Tokens (representing loans), DAI and USDC stablecoins, the Configurator (for protocol settings), Libraries (reusable modules), and Oracles (for asset pricing and interest rates).

AAVE's website offers a user-friendly interface where users can:

- **Deposit Assets**: Users can deposit various cryptocurrencies into liquidity pools, earning interest based on the demand for those assets.
- **Borrow Assets**: Users can borrow assets by providing collateral, with the interest rate determined by the asset's supply and demand.
- **Earn Interest**: Depositors earn interest on their assets, with rates that fluctuate based on market conditions.
- **Participate in Governance**: AAVE token holders can vote on protocol changes, influencing the direction and development of the platform.

AAVE token holders participate in the governance of the protocol through voting on various proposals. These proposals can include changes to protocol parameters, addition of new assets, and upgrades to the protocol. This decentralized governance model ensures that the community has a say in the direction of AAVE's development.

Technical Terms and Mechanisms

- **Collateral and Borrowing**:
  - **Liquidity Pools**: These are pools of assets provided by users, which can be borrowed by other users. Each liquidity pool supports various cryptocurrencies.
  - **Interest Rates**: Interest rates on AAVE are determined algorithmically based on supply and demand dynamics. When an asset's liquidity is low, the interest rate increases to attract more deposits.
  - **Collateralization Ratio**: This is the ratio of collateral value to loan value. AAVE requires borrowers to maintain a minimum collateralization ratio to secure their loans.
  - **Loan-to-Value (LTV) Ratio**: This ratio indicates the maximum amount a user can borrow against their collateral. For example, an LTV of 75% means a user can borrow up to 75% of the value of their collateral.
- **Liquidation Mechanisms**:
  - **Liquidation Threshold**: This is the percentage at which a loan is considered under-collateralized and subject to liquidation. If the value of the collateral falls below this threshold, a portion of the collateral is sold to repay the loan.
  - **Liquidation Penalty**: This is a fee applied when a liquidation occurs, intended to incentivize maintaining proper collateralization and compensating liquidators.

Formulas and Calculations

- **Interest Rate Calculation**:

$$\text{Interest Rate} = \text{Base Rate} + \left(\frac{\text{Utilization Rate}}{\text{Target Utilization Rate}}\right) \times \text{Variable Rate} \quad (1)$$

  Where:
  - **Base Rate** is the minimum interest rate.
  - **Utilization Rate** is the ratio of borrowed funds to total available funds in a pool.
  - **Target Utilization Rate** is a predetermined optimal utilization ratio.
  - **Variable Rate** adjusts based on the utilization rate.
- **Health Factor**:

$$\text{Health Factor} = \frac{\text{Collateral Value} \times \text{Liquidation Threshold}}{\text{Borrowed Amount}} \quad (2)$$

  A Health Factor above 1 indicates a safe loan, while below 1 indicates a risk of liquidation.

– **Liquidation Price**:

$$\text{Liquidation Price} = \frac{\text{Borrowed Amount}}{\text{Collateral Amount} \times \text{Liquidation Threshold}} \quad (3)$$

This formula calculates the price at which the collateral value will trigger liquidation.

Descriptives statistics

Processing on-chain data, such as that from AAVE, is complex due to the vast amount of data generated by transactions on the blockchain. To handle this effectively, we first used The Graph Platform. The Graph is a decentralized protocol for indexing and querying data from blockchains, starting with Ethereum. It allows developers to create "subgraphs" which define how blockchain data is aggregated and accessed. We typically use Messari's subgraph (see https://github.com/messari/subgraphs/blob/master/schema-lending.graphql). Initially, we created a notebook for querying. that produced raw files of monthly market data in JSON format, along with snapshots of positions, spanning from November 2020 to May 2023. Subsequently, we processed the obtained data, resulting in two datasets: 'market data' and 'position snapshot'. Table 1 outlines the primary components of our dataset.

Market Data	**Position Snapshot**
• Asset Price	• User Account ID
• Total Value Locked	• Position Identifier
• Total Deposits	• Asset Type
• Total Borrows	• Position Size
• Health Factor	• Position Value (USD)
• Interest Rates	• Position Type (e.g., Collateral)
• Reserve Metrics	• Timestamp
• Token Identifiers	• Risk Metrics

Table 1: Key Components of Market and Position Data

The market data, comprising 878,601 daily entries across 37 AAVE markets from December 2020 to December 2023, provides a macro view of protocol activity. It includes aggregated metrics such as token prices, Total Value Locked (TVL) , and various balance and rate indicators for Aave interest-bearing assets like DAI and WETH. Position snapshots, totaling 1,514,611 records over the same period, offer micro-level insights into individual user interactions. These entries include unique identifiers, asset information, balances, and position characteristics, enabling granular analysis of user behavior and risk exposure. This comprehensive dataset allows for in-depth analysis of AAVE's operational dynamics and user behavior over time, providing both broad market overviews and detailed individual position information.

The following section provides a detailed overview of the AAVE protocol by focusing on key aspects such as user composition, interest rates, market usage, and overall protocol health. The selected figures present crucial insights into the dynamics and structure of the AAVE ecosystem.

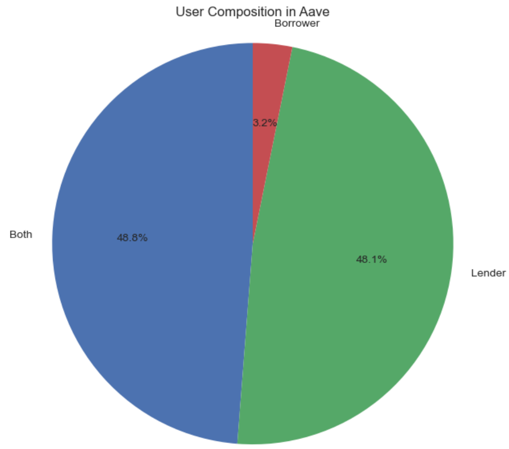

Figure 2: User Composition in AAVE: Distribution of Borrowers, Lenders, and Users Engaged in Both Activities.

Figure 2 shows the user composition within AAVE, where the majority of participants are either lenders (48.1%) or engaged in both lending and borrowing activities (48.8%). Borrowers alone constitute a smaller portion (3.2%), highlighting the protocol's focus on lending, with most users acting as liquidity providers or engaging in both sides of the market. This balance underscores the integrated nature of AAVE's user base, which is pivotal for the stability of its financial ecosystem.

Figure 3: Average Rates by Asset Category: Comparison of Stable Borrow Rates, Variable Borrow Rates, and Lend Rates across Different Asset Categories.

Interest rates within AAVE vary significantly by asset category, as depicted in Figure 3. Stablecoins exhibit notably higher borrow rates, particularly under variable conditions, while lending rates remain relatively lower across all asset types. These differences reflect the market's risk perception and demand, with stablecoins being more actively utilized, likely due to their lower volatility compared to other cryptocurrencies. Understanding these rate structures is essential for interpreting the financial incentives driving user behavior within the protocol.

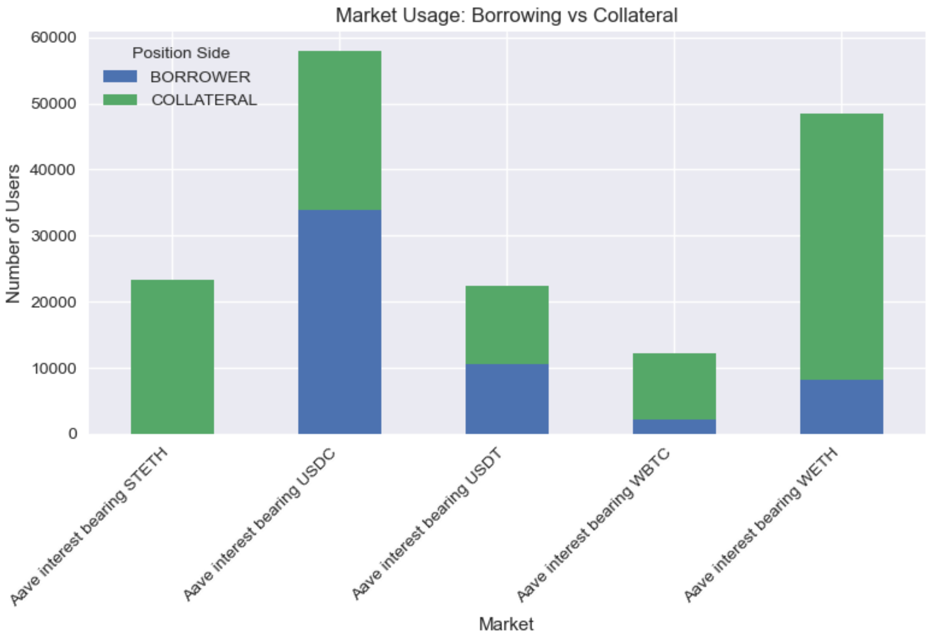

Figure 4: Market Usage: Borrowing vs Collateral: Number of Users Borrowing or Using Assets as Collateral Across Different AAVE Markets.

Figure 4 highlights the distribution of borrowing and collateral activities across various AAVE markets. Assets like USDC and WETH dominate both borrowing and collateral roles, signifying their central importance within the protocol. The significant engagement in these assets points to their reliability and trustworthiness in managing both lending and borrowing activities, making them critical for maintaining AAVE's liquidity and stability.

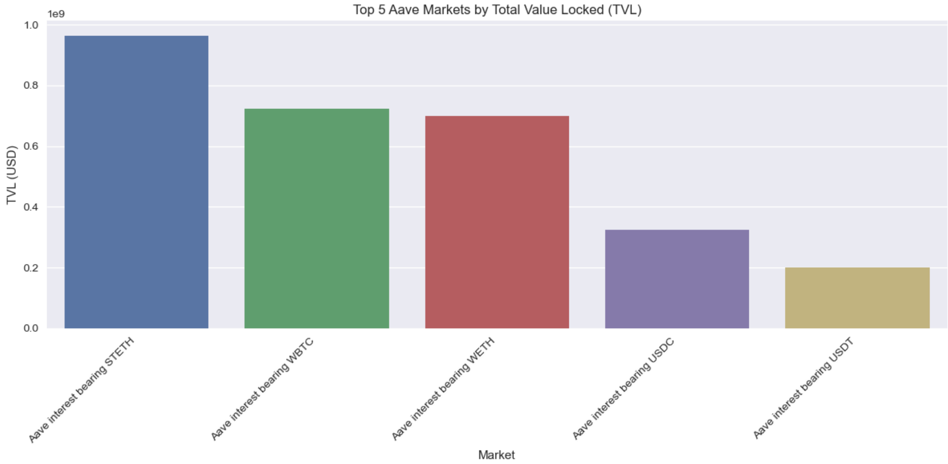

Figure 5: Top 5 AAVE Markets by Total Value Locked (TVL): Comparison of the Top 5 AAVE Interest-Bearing Markets Based on Total Value Locked in USD.

Total Value Locked (TVL) is a key indicator of a protocol's health, and Figure 5 presents the top five markets by TVL within AAVE. STETH, WBTC, and WETH are the most substantial, reflecting where user assets are most concentrated. This concentration of liquidity is a double-edged sword, providing robustness to these markets while also representing potential points of systemic risk should these assets experience significant volatility.

In summary, these descriptive statistics provide a comprehensive view of the AAVE protocol, highlighting the distribution of user roles, the structure of interest rates, the significance of key markets, and the temporal behavior of interest rates. Together, these factors create a nuanced understanding of AAVE's financial ecosystem, essential for any robust analysis or forecasting model focused on decentralized finance.

5 Network modelling

The analysis of AAVE's network structure is crucial for understanding the dynamics of liquidity, risk distribution, and the influence of key participants. By examining the network's topological properties and its evolution over time, we can gain insights into how network effects impact the protocol's stability and performance.

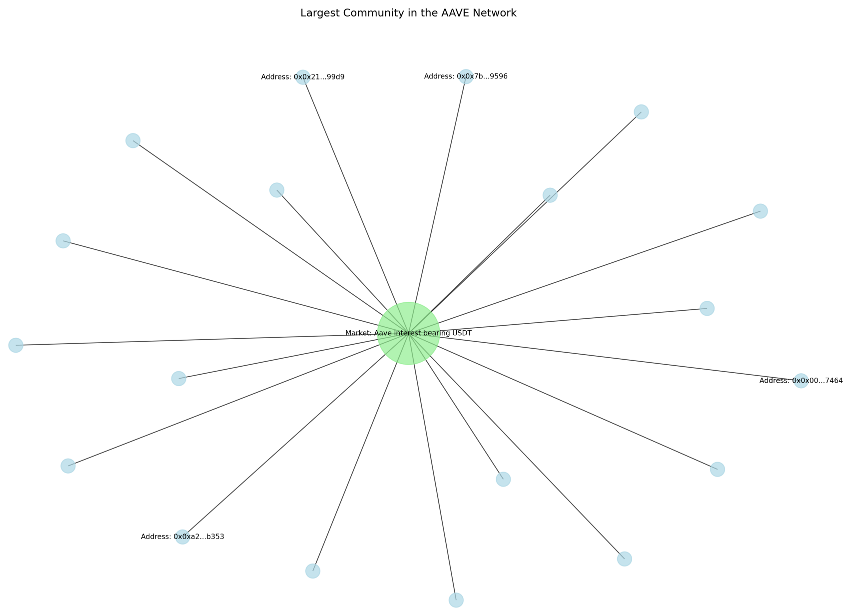

Figure 6: Largest Community in the AAVE Network: Visualization of the most significant cluster within the network, centered around the USDT market.

Figure 6 depicts the largest community within the AAVE network, highlighting its centrality around the USDT market. This cluster represents a critical hub in the network where significant lending and borrowing activities converge. The prominence of this community indicates the central role of stablecoins like USDT in the overall network, reinforcing their importance in maintaining liquidity and mitigating volatility.

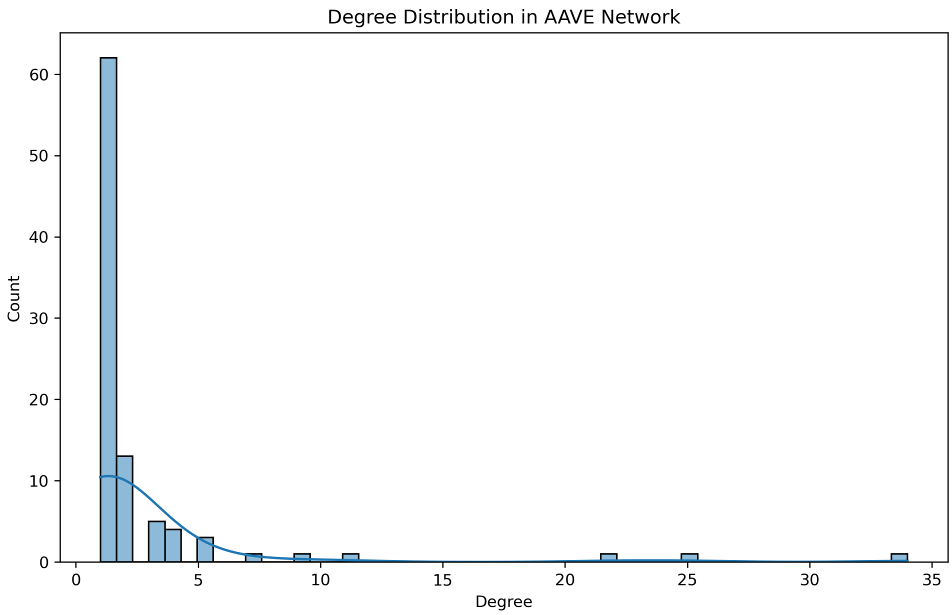

Figure 7: Degree Distribution in AAVE Network: Distribution of connections per node, emphasizing the presence of highly connected hubs.

The degree distribution in Figure 7 further underscores the presence of a few highly connected hubs within the network. These central nodes play a pivotal role in facilitating liquidity and ensuring the robustness of the protocol. The long-tail distribution reflects the network's scale-free nature, where most nodes have few connections, while a small number of nodes exhibit a high degree of connectivity, making them critical to the network's integrity and resilience.

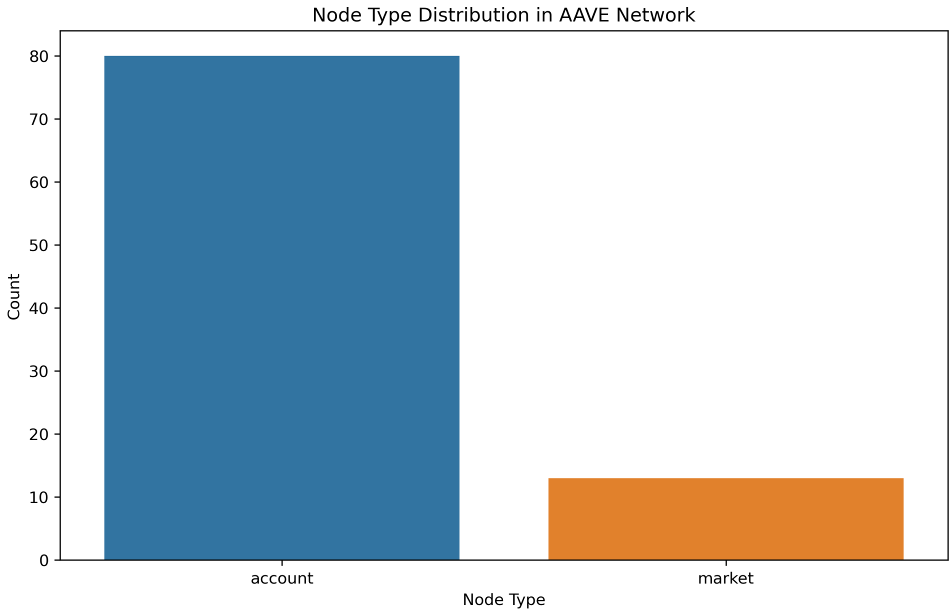

Figure 8: Node Type Distribution in AAVE Network: Comparison of account nodes versus market nodes.

Figure 8 illustrates the distribution of node types within the network, with account nodes significantly outnumbering market nodes. This imbalance highlights the dispersed nature of participation, where numerous users interact with a relatively smaller set of markets. Such a structure is typical in decentralized finance, where user interactions are heavily focused around key liquidity pools, reflecting both the concentration of capital and the distributed risk across the protocol.

PageRank is an algorithm originally used by Google to rank web pages in their search engine results. It assigns a numerical weighting to each element in a linked set, with the purpose of measuring its relative importance within the set. The PageRank of a page P_i is given by:

$$PR(P_i) = \frac{1-d}{N} + d \sum_{P_j \in M(P_i)} \frac{PR(P_j)}{L(P_j)}$$

where:

- d is the damping factor (usually set to 0.85),
- N is the total number of pages,
- $M(P_i)$ is the set of pages linking to P_i,
- $L(P_j)$ is the number of outbound links on page P_j.

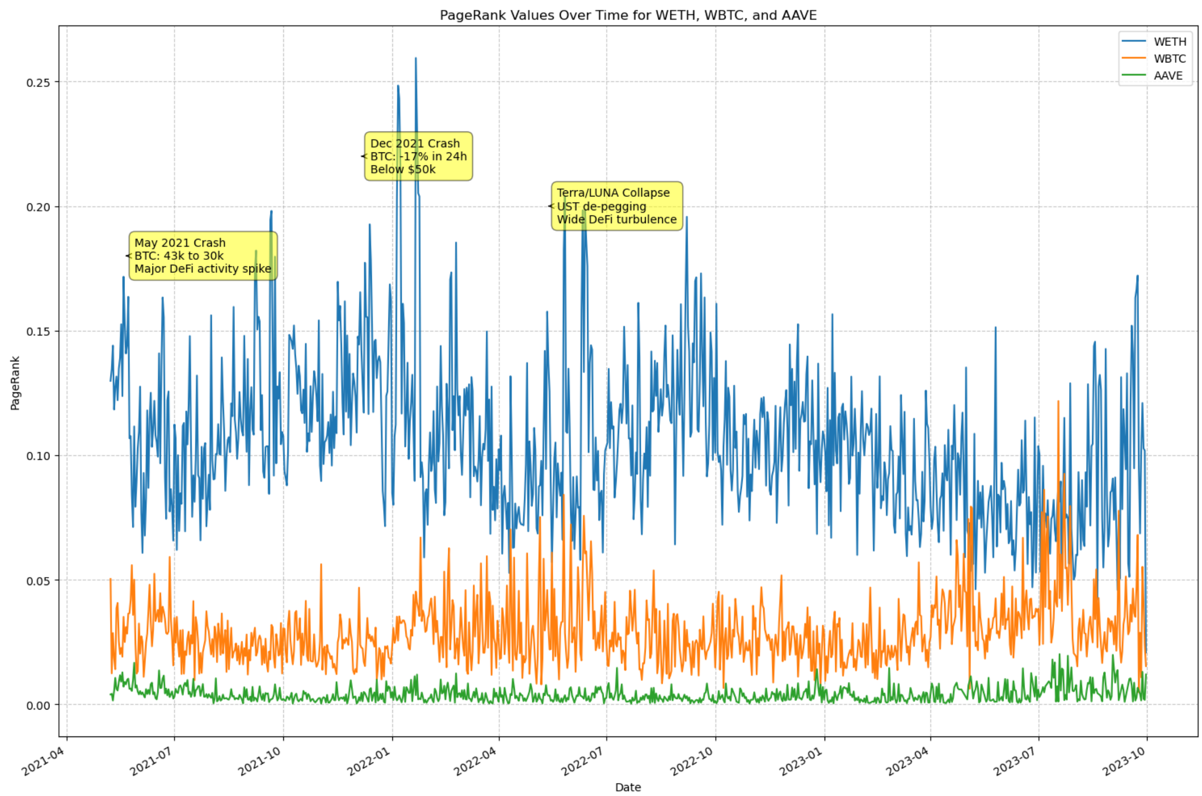

Figure 9: PageRank Values Over Time for WETH, WBTC, and AAVE. This plot illustrates the average PageRank values of WETH (blue), WBTC (orange), and AAVE (green) tokens from April 2021 to October 2023. Significant cryptocurrency market events, such as the May 2021 crash, December 2021 crash, and the May 2022 Terra/LUNA collapse, are annotated on the graph.

In the context of DeFi, PageRank helps identify key assets and users within the network that significantly impact price dynamics and systemic stability. By considering both direct and indirect interactions, PageRank can identify central nodes that act as major hubs of activity. These nodes, when impacted, can have cascading effects on the network, influencing liquidity of course but also tokens prices [32].

In our analysis, we computed two key metrics—*avg_pagerank* and *total_pagerank*—to measure the importance of DeFi tokens (WETH, WBTC, and AAVE) within the AAVE protocol. Using market and position data, we constructed daily network graphs where nodes represent user accounts and markets, with edges weighted by the USD value of positions. The PageRank algorithm was applied to these graphs to determine the relative importance of each market.

avg_pagerank is the average of the PageRank values for all markets associated with a token on a given day, reflecting the typical importance of the token's markets. It is calculated as:

$$\text{avg\_pagerank}(T) = \frac{1}{n}\sum_{i=1}^{n} PR(market_i)$$

total_pagerank is the sum of these PageRank values, indicating the overall influence of the token across the network on that day. It is given by:

$$\text{total\_pagerank}(T) = \sum_{i=1}^{n} PR(market_i)$$

These metrics help us understand the evolving significance of these tokens in the DeFi ecosystem, especially during periods of market turbulence.

Figure 9 captures the historical changes in the importance of three major DeFi tokens (WETH, WBTC, and AAVE) over a period marked by significant market turbulence. Using PageRank as a measure, the analysis focuses on how market events, such as major price crashes and the collapse of Terra/LUNA, impacted the DeFi landscape.

During the May 2021 crash, when Bitcoin (BTC) fell from 43k to 30k, there was a noticeable spike in the PageRank value of WETH, indicating increased significance and activity in the DeFi space, likely due to trading and liquidity movements triggered by the volatility. The December 2021 crash, which saw BTC drop 17% within 24 hours, led to another peak in WETH's PageRank, highlighting its crucial role during this period of market stress. The May 2022 Terra/LUNA collapse, resulting in the de-pegging of UST and widespread DeFi turbulence, caused sharp fluctuations in the PageRank values of WETH and WBTC, reflecting their central role as the market adjusted rapidly. These events underscore the strong correlation between market disruptions and the prominence of these tokens in DeFi protocols, as evidenced by their rising PageRank values.

In summary, the network analysis reveals a structure characterized by central hubs and a distribution of risk across a wide base of participants. The stability of the AAVE protocol hinges on these network dynamics, where a few key nodes maintain liquidity and ensure the system's robustness. Understanding these network properties is essential for predicting how changes in market conditions might propagate through the system, potentially impacting the price dynamics of key assets.

6 Tokens price forecasting

The objectiv is to assess the impact of integrating PageRank features into an LSTM model for forecasting the prices of two tokens, WETH and WBTC. To implement this approach, we construct a network representation of the AAVE ecosystem, where nodes represent assets, users, and smart contracts, and edges represent interactions such as lending, borrowing, and collateralization. The PageRank algorithm is then applied to this network to rank nodes based on their centrality and influence.

Table 2: Root Mean Squared Error (RMSE) for Different Models Applied to Bitcoin Data

Model	**RMSE**
LSTM	4122.40
Linear Regression	6586.71
Random Forest	5450.45
GRU	4264.11

We sought to determine whether Long Short-Term Memory (LSTM) networks indeed outperform other models, as indicated in the existing literature. We conducted feature selection, focusing on key attributes relevant to our data set, and compared the performance of several models: LSTM, Linear Regression, Random Forest, and Gated Recurrent Unit

(GRU). The results, summarized in Table 2, show the Root Mean Squared Error (RMSE) for each model applied to Bitcoin data:

- LSTM: 4122.40
- Linear Regression: 6586.71
- Random Forest: 5450.45
- GRU: 4264.11

The LSTM model achieved the lowest RMSE, confirming our hypothesis and validating our decision to utilize LSTM for this analysis. To clarify, RMSE is a standard metric for assessing the accuracy of predictive models. It is calculated by taking the square root of the average of the squared differences between actual and predicted values. A lower RMSE indicates better predictive performance.

6.1 Long Short-Term Memory (LSTM) Networks

Long Short-Term Memory (LSTM) networks are a specialized form of Recurrent Neural Networks (RNNs) designed to overcome the limitations of traditional RNNs, particularly the issues of vanishing and exploding gradients. The strength of LSTMs lies in their ability to learn long-term dependencies in sequential data by maintaining a cell state, which is controlled through three key gates: the input gate, forget gate, and output gate. Below, we delve into the mathematical formulation of LSTMs and their relevance to cryptocurrency price prediction.

An LSTM unit is defined by the following components:

- **Cell State (C_t)**: The cell state is a memory that carries information across the sequence. It is updated by the forget and input gates at each time step t.
- **Forget Gate (f_t)**: The forget gate determines what portion of the previous cell state (C_{t-1}) should be retained. It is defined as:

  $$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

  where:

  - σ is the sigmoid activation function.
  - W_f and b_f are the weight matrix and bias for the forget gate.
  - h_{t-1} is the hidden state from the previous time step.
  - x_t is the input at the current time step.

- **Input Gate (i_t) and Candidate Cell State ($\tilde{C}_t$)**: The input gate controls how much of the new information will be added to the cell state. The candidate cell state represents new potential values that could be added to the cell state:

  $$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

  $$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

  where:

- W_i, W_C, b_i, and b_C are the weight matrices and biases for the input gate and the candidate cell state, respectively.
- tanh is the hyperbolic tangent activation function.

- **Cell State Update**: The cell state C_t is updated using the forget gate and input gate as follows:
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$
This equation ensures that the LSTM can retain information over long sequences if necessary, or discard irrelevant information.

- **Output Gate (o_t) and Hidden State (h_t)**: The output gate determines the next hidden state, which is also the output of the LSTM cell at time step t:
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \cdot \tanh(C_t)$$
where W_o and b_o are the weight matrix and bias for the output gate.

The combination of these gates and their respective operations allows the LSTM to control the flow of information, selectively remembering or forgetting parts of the input sequence, which is crucial for learning long-term dependencies in sequential data. Cryptocurrency price prediction is inherently a time series problem, where past price movements influence future prices. The challenge lies in accurately capturing both short-term fluctuations and long-term trends. The LSTM's architecture is particularly well-suited for this task due to its ability to maintain long-term dependencies through its cell state mechanism. Given the sequential nature of price data, where patterns and dependencies exist across different time horizons, the mathematical structure of LSTMs enables the model to learn complex temporal dynamics without suffering from the vanishing gradient problem that hampers traditional RNNs. The gates in an LSTM ensure that relevant information from previous time steps is preserved while irrelevant information is discarded, allowing the model to focus on meaningful patterns that are crucial for accurate prediction.

6.2 Data processing

The dataset underwent several transformations to ensure it was ready for time-series forecasting with LSTM. First, the features *balanceUSD*, *totalValueLockedUSD*, *totalBorrowBalanceUSD*, *variableBorrowRate*, *stableBorrowRate*, *supplyRate*, *avg_pagerank*, and *total_pagerank* were lagged by 1, 2, and 3 days to capture relevant historical information. After generating these lagged features, the original columns representing the current day (t) were removed to prevent data leakage. Finally, the dataset was further prepared for LSTM by excluding non-numeric columns such as *timestamp*, *token*, *id*, and *date*.

We integrated two key PageRank metrics computed before, *avg_pagerank* and *total_pagerank*, to evaluate the importance of specific DeFi tokens (WETH, WBTC, AAVE) within the AAVE protocol. The *avg_pagerank* reflects the average importance of all markets associated with a token on a given day, while the *total_pagerank* represents the cumulative influence of the token across the network on that day. These metrics provide insights into the evolving role of these tokens within the DeFi ecosystem, particularly during periods of market stress.

During data preprocessing, we also considered the inclusion and exclusion of specific columns to prevent issues such as data leakage, multicollinearity, and overfitting. Data leakage occurs when information from the future influences the model, leading to unrealistically high performance during training but poor generalization to new data. To avoid this, columns that include future information are excluded. Additionally, multicollinearity, where highly correlated features can distort model coefficients, was addressed by removing redundant columns. This ensures that the model remains interpretable and reduces the risk of overfitting, where the model performs well on the training data but fails to generalize to unseen data.

We end up with WETH dataset in two versions: one with PageRank metrics and one without. The dataset with PageRank includes columns such as 'timestamp', 'balanceUSD', 'inputTokenPriceUSD', 'totalValueLockedUSD', various interest rates, and the computed PageRank metrics ('avgpagerank' and 'totalpagerank'). The full WETH dataset, while similar, excludes the PageRank metrics and focuses purely on financial data and token-specific metrics.

Similarly, the WBTC dataset is also available in two formats. The version with PageRank and without mirror the structure of the WETH datasets.

6.3 Predictions

Following the data preparation, both the feature set (X) and the target variable (y) undergo scaling using MinMaxScaler. This scaling step is crucial as it normalizes the input range, facilitating faster convergence of the LSTM model during training. The dataset is then split into training and testing subsets in an 80-20 ratio, with no shuffling, thereby preserving the inherent temporal sequence of the data.

The architecture of the LSTM model is designed to capture and learn temporal dependencies and relationships within the data. The model comprises two LSTM layers followed by two Dense layers. The sequential nature of the LSTM layers allows the model to effectively process time series data, while the Dense layers help refine the final output.

Training of the model is conducted using the Adam optimizer and the mean squared error loss function. The model is trained over 10 epochs with a batch size of 1, a configuration selected to maximize the model's ability to learn from the time-dependent data. Post-training, the model's performance is rigorously evaluated on the test data. Key metrics used for evaluation include RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and R^2. These metrics provide a comprehensive assessment of the model's predictive accuracy and its ability to generalize to unseen data.

The performance of our LSTM models for predicting the prices of WETH and WBTC was evaluated both with and without incorporating PageRank features. The models were trained using a dataset processed to include lagged features to capture historical trends and dependencies, with the PageRank features providing additional context on the tokens' importance within the DeFi ecosystem.

- **Performance Metrics:** The models were evaluated using the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The RMSE is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

where y_i represents the actual price, $\hat{y}_i$ represents the predicted price, and n is the number of data points. A lower RMSE indicates better predictive accuracy.

The MAPE is calculated using the following formula:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{5}$$

where y_i is the actual price, $\hat{y}_i$ is the predicted price, and n is the number of data points. MAPE expresses the accuracy as a percentage, with lower values indicating better predictive performance.

- **Results:** The table below summarizes the RMSE and MAPE for both tokens:

Table 3: Comparison of Model Results With and Without PageRank Feature for WETH and WBTC

Token	Model	RMSE	MAPE
WETH	Without PageRank	216.053	11.2942%
	With PageRank	130.156	6.1716%
WBTC	Without PageRank	1897.435	13.5682%
	With PageRank	843.276	6.4129%

- **Analysis:**
  - For **WBTC**, incorporating PageRank features resulted in lower RMSE and MAPE values, suggesting that the enhanced model with PageRank features provides a more accurate prediction of WBTC prices.
  - For **WETH**, the model with PageRank features also produced lower RMSE and MAPE values compared to the model without these features. This indicates that the inclusion of PageRank features improves the predictive accuracy for WETH as well.
  - The inclusion of PageRank features can improve the accuracy of price predictions for both assets (e.g., WBTC and WETH), suggesting that the effectiveness of network-based features like PageRank may be advantageous across different tokens. Further investigation is encouraged to refine the use of such features in predictive modeling.

7 Conclusion

The analysis conducted in this study reveals that incorporating network-based approaches, such as PageRank, into machine learning models can significantly enhance predictive accuracy for certain DeFi tokens. For instance, the results indicate that both WBTC and WETH benefited from the inclusion of PageRank features, as evidenced by the lower RMSE and MAPE values achieved with these models.However, the findings also underscore the complexity of DeFi ecosystems, where the dynamics of token prices are influenced not only by direct financial metrics but also by broader network interactions within the protocol. While the enhanced models with PageRank features provided more accurate predictions for both WBTC and WETH, it is important to note that the performance improvement may not be uniform across all contexts. The results highlight the potential for network features to improve forecasting models, but they also indicate a need for careful refinement in their application. The effectiveness of such features appears to be asset-specific; while they improved accuracy for both WBTC and WETH in this study, different assets may react differently to network-based features.

Additionally, the analysis may be limited by the way PageRank features were integrated, potentially without specific tuning for price prediction tasks. Centrality metrics like PageRank capture network importance, but their relevance to price prediction can vary depending on the context and behavior of the asset in question. For WETH, the improvements were marked, suggesting that the incorporation of network features may effectively enhance model performance when applied appropriately. Future improvements could involve more sophisticated feature engineering, such as integrating PageRank with other centrality measures or applying temporal adjustments to better capture the dynamics relevant to price changes. Experimenting with more granular or context-specific PageRank calculations may further enhance the models' ability to predict price movements accurately. Overall, this study contributes to a deeper understanding of DeFi markets, particularly in the context of decentralized lending, and provides a foundation for developing more targeted strategies in price forecasting and risk management within these innovative financial ecosystems.

References

[1] Harvey et al., 2021. Impact of social metrics in decentralized finance.

[2] Schär, 2021. Impact of social metrics in decentralized finance.

[3] Whitepaper.io, 2020. [Online]. Available: https://whitepaper.io/coin/aave.

[4] Coinmarketcap, 2022. [Online]. Available: https://coinmarketcap.com/currencies/aave/.

[5] Gudgeon, Lewis, Perez, Daniel, Harz, Dominik, Livshits, Benjamin, & Gervais, Arthur, 2020. The Decentralized Financial Crisis, pp. 1-15. DOI: 10.1109/CVCBT50464.2020.00005.

[6] Bartoletti, M., Chiang, J.Hy., Lafuente, A.L. (2021). SoK: Lending Pools in Decentralized Finance. In: Bernhard, M., et al., Financial Cryptography and Data Security. FC 2021 International Workshops. FC 2021. Lecture Notes in Computer Science, vol 12676. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-662-63958-0_40.

[7] Diamond, Douglas W., & Dybvig, Philip H., 1983. Bank Runs, Deposit Insurance, and Liquidity.

[8] Eisenberg, Laurence K., & Noe, Thomas H., 2001. Systemic Risk in Financial Networks. Management Science, Vol. 47, No. 2, pp. 236-249. Available at SSRN: https://ssrn.com/abstract=993629 or http://dx.doi.org/10.2139/ssrn.173249.

[9] Tovanich, Natkamon, Kassoul, Myriam, Weidenholzer, Simon, & Prat, Julien, 2023. Contagion in Decentralized Lending Protocols: A Case Study of Compound. Proceedings of the 2023 Workshop on Decentralized Finance and Security, ACM, Nov 2023, Copenhagen, Denmark. DOI: ff10.1145/3605768.3623544ff.

[10] Lin, L., Liang, D. (2022). Blockchain and DeFi: The Rise of Decentralized Financial Systems. IEEE Engineering Management Review, 50(1), 44-54.

[11] Qin, K., Zhou, L., Gervais, A. (2021). Quantifying Blockchain Extractable Value: How dark is the forest?. arXiv preprint arXiv:2101.05511.

[12] Antoni, M., Gu, Q., Sornette, D. (2021). Generalized Leverage Effect in DeFi: Evidence from Compound and Aave. arXiv preprint arXiv:2106.12460.

[13] Pereira, J., Tavalaei, M. M., Ozalp, H. (2022). Blockchain-based platforms: Decentralized infrastructures and its boundary conditions. Technological Forecasting and Social Change, 175, 121339.

[14] Liu, Z., Luong, N. C., Wang, W., Niyato, D., Wang, P., Liang, Y. C., Kim, D. I. (2022). A survey on blockchain: A game theoretical perspective. IEEE Access, 7, 47615-47643.

[15] Angeris, G., Chitra, T. (2020). Improved price oracles: Constant function market makers. In Proceedings of the 2nd ACM Conference on Advances in Financial Technologies (pp. 80-91).

[16] Castro-Iragorri, C., Ramírez, J., & Vélez, S. (2021). Financial intermediation and risk in decentralized lending protocols. Banking & Insurance eJournal.

[17] Clements, R. (2021). Emerging Canadian Crypto-Asset Jurisdictional Uncertainties and Regulatory Gaps. CGN: Governance Law & Arrangements by Subject Matter (Topic).

[18] Tariq, S. A. S., Iftikhar, S., Iftikhar, K., Raza, H., & Idrees, S. (2023). The Role of Digital Finance in Economic Development: A Cross Country Analysis. Journal of Policy Research.

[19] Kolyandov, S. (2021). The Rising Popularity of Digital Transaction Platforms. Trakia Journal of Sciences.

[20] Qin, Kaihua, Zhou, Liyi, Livshits, Benjamin, and Gervais, Arthur, 2021. An Empirical Study of DeFi Liquidations: Incentives, Risks, and Instabilities. Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21). DOI: 10.1145/3460120.3484800.

[21] Saengchote, K. 2023. Leverage and Liquidity Risks in Decentralized Finance Platforms. Journal of Financial Stability, 60, 101010. DOI: 10.1016/j.jfs.2023.101010.

[22] Altman, E. I., 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. Journal of Finance, 23(4), 589-609.

[23] Ohlson, J. A., 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research, 18(1), 109-131.

[24] Barboza, F., Kimura, H., & Altman, E. I., 2017. Machine Learning Models and Bankruptcy Prediction. Expert Systems with Applications, 83, 405-417.

[25] Chen, H., & Du, Y., 2020. Artificial Intelligence in Financial Market Prediction. Journal of Financial Markets, 50, 100-115.

[26] Zhao, H., & Zhao, H., 2023. Predicting Financial Distress using Machine Learning Techniques. Applied Soft Computing, 114, 107-118.

[27] Werner, S. M., Perez, D., & Gudgeon, L. (2021). Sok: Decentralized finance (defi). *arXiv preprint arXiv:2101.08778.*

[28] Zeng, J., & Chen, H. (2021). Blockchain-based decentralized finance: The infrastructure and data analysis. *Journal of Financial Data Science*, 3(1), 1-20.

[29] Bartoletti, M., Chiang, J. H., & Lluch-Lafuente, A. (2021). A formal model of liquidation in DeFi lending. *International Conference on Blockchain and Cryptocurrency (ICBC)*, 1-10.

[30] Chen, J., & Zhao, Y. (2021). Deep learning for cryptocurrency price prediction: A comprehensive review. *Journal of Financial Data Science.*

[31] Jiang, X., & Zhang, H. (2021). Bitcoin price prediction using machine learning algorithms and attention mechanisms. *Financial Innovation.*

[32] Lin, W., & Liang, P. (2022). Network-based analysis of DeFi token price stability: A case study. *Blockchain Research and Applications.*

[33] Wu, L., & Liu, J. (2022). Hybrid models for DeFi price prediction: Integrating network analysis and machine learning. *Journal of Financial Markets.*

[34] Balcilar, M., & Demirer, R. (2021). Sentiment analysis for DeFi token price prediction using Twitter data. *Journal of Computational Finance.*