**Q1 Data Understanding**

i. The paper mainly uses historical stock market data, specifically daily prices including opening price, closing price, high price, low price and trading volume.

From this historical data, various technical indicators are computed to capture trends and patterns. Examples of technical indicators used include moving averages, exponential moving average, relative strength index, moving average convergence divergence, stochastic oscillator and on-balance volume (Mejia et al.)

ii. Importance of using such indicators in forecasting stock price trends:

The transformation of raw market data through mathematical methods in technical indicators enables analysts to extract meaningful patterns which reveal market momentum and volatility and trend strength. Technical indicators transform historical price and volume data through mathematical operations to reveal patterns which are concealed within noisy financial time series.

These indicators provide several advantages:

* Moving averages and similar indicators smooth market fluctuations to help investors identify bullish or bearish trends in the market.
* The RSI and Stochastic indicators along with other oscillators enable users to measure market momentum while detecting potential price reversals.
* Bollinger Bands serve as a tool which allows models to determine market risk levels and price variability according to (Murphy, 45).

The application of technical indicators enables machine learning models to extract vital patterns from noisy inputs which leads to better forecasting accuracy. The use of processed features obtained from technical indicators leads to better generalisation and financial interpretation of model predictions compared to using raw prices.

**Q2. Security Understanding: IVV (iShares Core S&P 500 ETF)**

**Fund overview:**

The iShares Core S&P 500 ETF (ticker: IVV) is a passively managed exchange-traded fund (ETF) designed to track the performance of the S&P 500 Index, representing 500 of the largest publicly traded U.S. companies. Managed by BlackRock, IVV provides investors with broad exposure to the U.S. large-cap equity market, covering sectors like technology, healthcare, financials, and consumer discretionary.

**Asset type:**

* Type: Equity ETF
* Underlying Index: S&P 500 Index
* Holdings: Approximately 500 large-cap U.S. stocks
* Top Holdings: Apple, Microsoft, Amason, NVIDIA, Alphabet

**Historical price performance:**

Since its inception in **May 2000**, IVV has demonstrated strong long-term performance, mirroring the historical growth trajectory of the S&P 500.

* 2000 – 2010: Modest growth with significant volatility due to the Dot-com crash and 2008 Financial Crisis.
* 2010 – 2020: Strong bull market performance, fuelled by technology sector growth.
* 2020 – Present: Recovery from the COVID-19 market crash, followed by volatility from interest rate hikes in 2022-2023.

**Key historical statistics:**

* Annualised 10-Year Return (as of 2024): ~11%
* Expense Ratio: 0.03% (one of the lowest among ETFs)
* Dividend Yield: Approximately 1.4%
* Assets Under Management (AUM): Over $300 billion

**Price snapshot:**

* All-time High: Around $480 (early 2022)
* Recent Price (April 2025): Approximately $450
* 52-week Range: $390–$460

**Other notable features:**

* Liquidity: Highly liquid with narrow bid-ask spreads.
* Suitability: Often used by long-term investors seeking exposure to the overall U.S. economy with a low-cost, diversified vehicle.
* Risk: Market risk tied to the broader performance of the U.S. economy and large-cap stocks.

**Classification versus Regression**

The authors selected classification instead of regression because they wanted to predict stock market direction between rise and fall rather than future price values (Mejia et al., 9). The modeling task becomes easier through classification because it transforms into a simple binary decision between rise and fall which traders find more useful. The evaluation becomes simpler with accuracy and F1-score instead of RMSE and the model complexity decreases while preventing overfitting to price fluctuations' noise.

The authors could have defined their classification variable through two different approaches:

**Threshold-based movement:** The classification should focus on the magnitude of changes instead of any movement. For example:

* The model should classify the day as up when daily return exceeds +0.5%.
* The model should classify days with returns below -0.5% as down.
* The model should classify days with returns between -0.5% and +0.5% as neutral.

**Volatility-adjusted classes:** The classification system should use volatility bands to determine its categories.

* Standard deviation of past returns should determine what constitutes significant market movement.
* The model should classify returns as significant up when they exceed one standard deviation above the mean.
* The model should classify returns as significant down when they fall below one standard deviation below the mean.
* The remaining cases should be classified as no significant change.

**Q3 Section 2: Data**

2.1 Data collection

2.2 Technical indicator calculation

2.3 Data processing

2.4 Feature selection

2.5 Classification variable definition

**Section 3: Methodology**

3.1 Neural network model

3.2 Hyperparameter optimisation with Genetic Algorithm

3.3 Feature selection using LASSO

3.4 Training and testing procedure

**Table 1.0: Dividing Descriptive statistics from Models**

|  |  |  |
| --- | --- | --- |
| **Aspect** | **Descriptive statistics** | **Predictive Models** |
| Purpose | Understand the structure, distribution, and relationships in data | Predict future outcomes or classify new data |
| Example | Pearson Correlation measures the linear relationship between two variables | LASSO Regression selects important predictors and builds a model by shrinking coefficients toward zero for irrelevant features |
| Process | Summarises data | Learns from data |
| Outputs | Correlations, means, variances, graphs | Predictions, selected features, coefficients |
| Dependency on Labels | No | Yes |

**Optimisation process of technical indicators**

The authors applied a two-step optimisation strategy to maximise the predictive ability of technical indicators:

Genetic algorithm optimisation

* A Genetic Algorithm is first used to optimise the parameter settings of each technical indicator, such as choosing the most effective time windows for moving averages or the RSI (Mejia et al., 6).
* The GA evolves populations over generations, selecting parameter combinations that yield the highest predictive performance (Mejia et al., 6).

LASSO feature selection

* After parameter optimisation, the authors use LASSO regression to automatically select the best subset of indicators by penalising less useful ones, thus reducing dimensionality (Mejia et al., 7).

**How the authors improve predictive power**

The combination of optimised parameters and feature selection enhances predictive power significantly:

* Customisation of Indicators: Fine-tuning standard technical indicators to fit the specific dynamics of the emerging markets improves signal quality.
* Noise Reduction: LASSO shrinks noisy or redundant inputs, reducing overfitting and focusing the neural network on truly informative features.
* Model Generalisation: By reducing the number of irrelevant features, the model becomes more robust and performs better on unseen data.

**Why It Is Important to Optimise Indicators for Neural Networks**

Optimising technical indicators is crucial because:

* Neural networks are sensitive to irrelevant features, and too many noisy inputs can lead to overfitting (Goodfellow et al., 97).
* Better feature quality means faster training and higher prediction accuracy, improving both efficiency and effectiveness (Hastie et al., 217).
* Dimensionality reduction lowers computational costs, enabling faster and more stable learning even on large financial datasets (Mejia et al., 7).

**References**

Murphy, John J. Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance, 1999.

Sagaceta Mejia, Gilberto, Ivan Marsura, Sergio Romero-Torres, and Edgar Sucar. “An Intelligent Approach for Predicting Stock Market Movements in Emerging Markets Using Optimised Technical Indicators and Neural Networks.” The Open Economics Journal, vol. 15, no. 1, 2022, <https://www.degruyter.com/document/doi/10.1515/econ-2022-0073/html>.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed., Springer, 2009.