**1. Stock Selection and Feature Engineering**

* **Objective**: Predict next-day returns for each stock using historical price-based indicators.
* **Technical Indicators Used**:
  + **SMA (Simple Moving Average)**: Average price over a window (e.g., 20 days).
  + **EMA (Exponential Moving Average)**: Similar to SMA but gives more weight to recent prices.
  + **RSI (Relative Strength Index)**: Measures recent gains vs. losses to assess momentum.
  + **Bollinger Bands (Upper, Middle, Lower)**: Gauge price volatility and identify overbought/oversold conditions.
  + **MACD (Moving Average Convergence Divergence)** and its **signal/histogram**: Used to capture trend strength and direction.

**2. Return Prediction Model**

* **Model Used**: Feedforward Neural Network (3 layers with dropout).
* **Inputs**: Scaled technical indicators.
* **Target**: One-day-ahead percentage return.
* **Training Frequency**: Every year using only past data — ensures realistic, non-biased training.
* **Output**: A predicted return for each stock, used in portfolio optimization.

**3. Portfolio Construction**

* **Expected Return**: Calculated as the average predicted return per stock.
* **Risk Estimation**: Based on historical **covariance matrix** of stock returns over the past 252 days.
* **Optimization Objective**: Maximize **Sharpe Ratio**, which balances return against risk.
* **Constraints**:
  + Total weights must sum to 1 (fully invested).
  + Maximum stock weight capped at 20% to reduce concentration risk.

**4. Risk Management**

**Key Techniques:**

* **Weight Truncation**: Limits any one stock’s weight to 20%, encouraging diversification.
* **Rebalancing Frequency**: Annual rebalancing minimizes overfitting and transaction costs.
* **Volatility Modeling**: Uses historical return data to account for risk via the covariance matrix.
* **Transaction Costs**: A 0.1% cost is factored in for every rebalance, simulating real-world frictions.
* **Drawdown Analysis**:
  + **Drawdown**: Measures how much the portfolio falls from its peak.
  + **Max Drawdown**: Largest observed drawdown; helps assess worst-case loss.

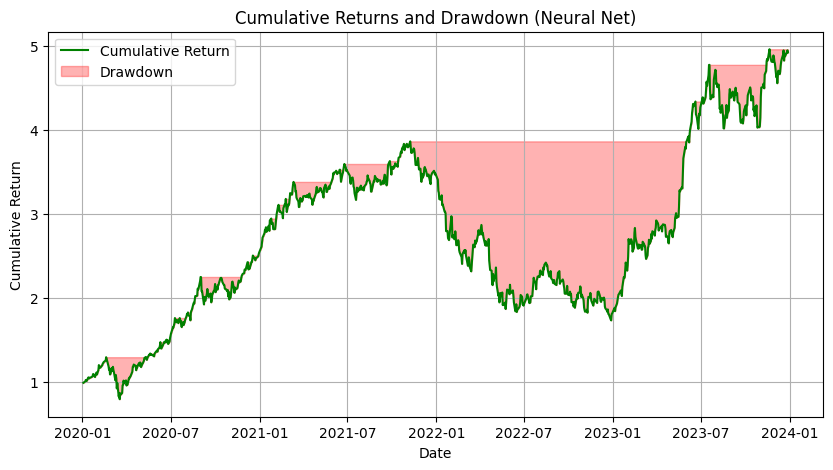
**5. Backtesting and Evaluation**

**Approach:**

* **Backtest Period**: 2020 to 2024.
* **Lookback Window**: 252 trading days (~1 year) of historical data used to compute volatility.
* **Evaluation Method**:
  + Train model on pre-2020 data.
  + Predict returns and optimize the portfolio each year.
  + Track how the portfolio performs over the next year.

**Performance Metrics:**

* **Mean Squared Error (MSE)**: Measures how far the model’s predictions are from actual returns. Lower is better.
* **R² (R-squared)**: Indicates how much variance in returns the model explains. Closer to 1 is better.
* **Cumulative Returns**: Tracks total growth of portfolio over time.
* **Annualized Return**: Compounded yearly growth of portfolio.
* **Annualized Volatility**: Standard deviation of returns scaled to a yearly basis.
* **Sharpe Ratio**: (Portfolio return - Risk-free rate) / Volatility — measures return per unit of risk.
* **Maximum Drawdown**: Largest drop from a historical peak to a trough — indicates downside risk.

Explanation of the Cumulative returns vs Drawdown graph

**Overview:**

This graph visualizes the performance of the neural network-based portfolio strategy over the backtesting period (**2020 to 2024**). It presents both the **cumulative return** of the portfolio and the corresponding **drawdowns**.

**Key Elements:**

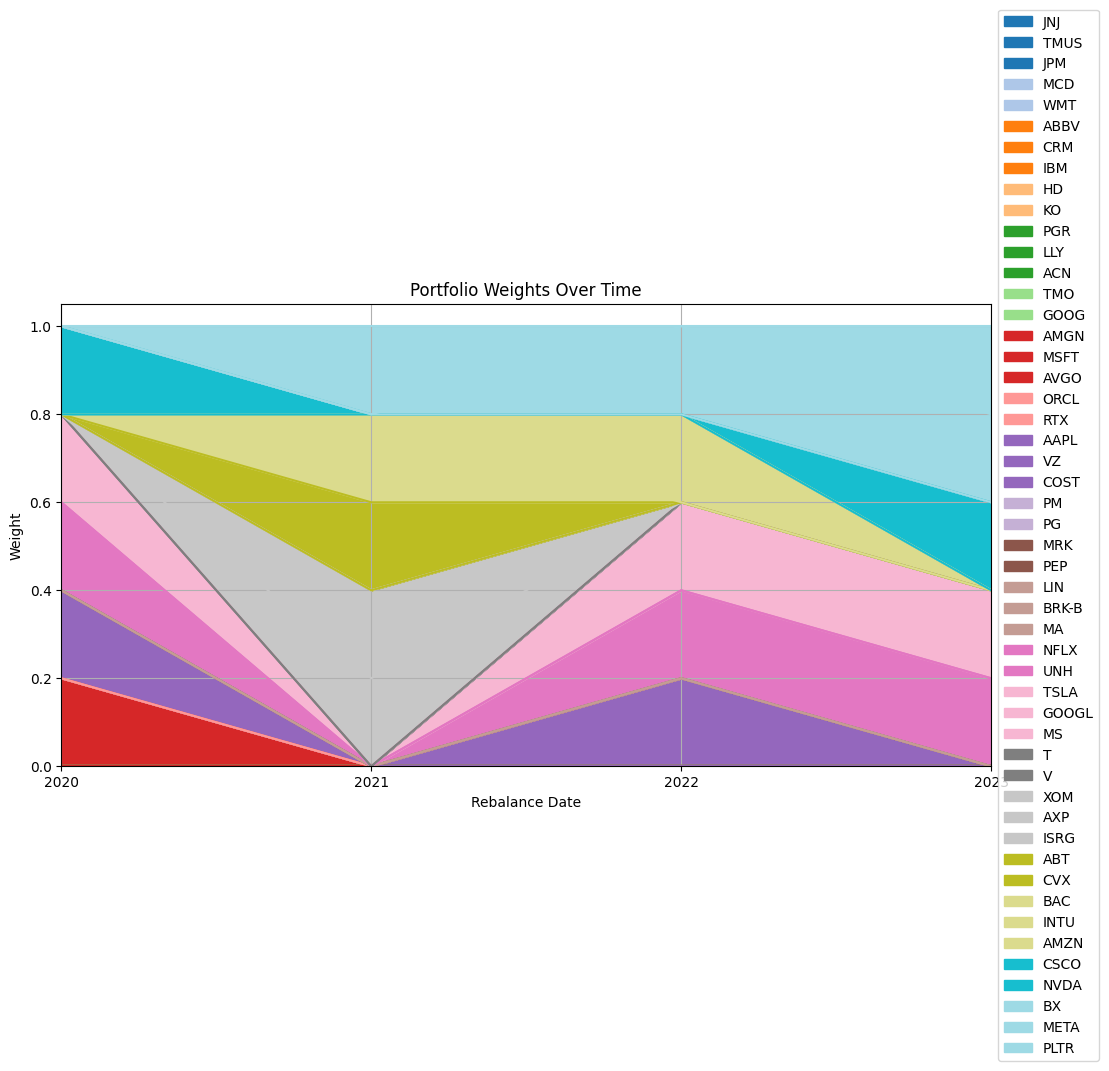
* **Green Line – Cumulative Return**:
  + Represents the **total growth** of the portfolio if an investor had followed the strategy from the start of 2020.
  + A value of 5.0 at the end of the period means the portfolio has grown by **400%** (i.e., 5× the initial investment).
* **Red Shaded Area – Drawdown**:
  + Highlights the **percentage drop** from the peak value of the portfolio at any point in time.
  + The deeper and wider the red area, the more **severe and prolonged** the loss period.
  + This is a key risk metric showing how much the portfolio declined before recovering.

**Interpretation:**

* From **early 2020 to late 2021**, the strategy shows a **steady upward trend**, reflecting strong model performance.
* Between **early 2022 and early 2023**, there is a **significant drawdown** — the portfolio experienced a major decline, illustrating market risk or model underperformance during that period.
* In **2023**, the strategy **recovers strongly**, reaching new highs by the end of 2023.
* Overall, despite temporary losses, the strategy achieved a **high cumulative return** with eventual recovery, showcasing both its **return potential** and **resilience**.

**Conclusion:**

This chart effectively demonstrates the trade-off between return and risk. While the neural network model can capture profitable trends, it is also subject to periods of drawdown, highlighting the importance of robust risk management in any active trading strategy.



**Overview:**

This **stacked area chart** visualizes how the **portfolio composition** (i.e., allocation of weights to different stocks) evolves over time during each **annual rebalancing period** from **2020 to 2024**.

**Key Elements:**

* **X-axis – Rebalance Date**:
  + Indicates each annual rebalancing point when the portfolio is reconstructed based on the updated predictions and market data.
* **Y-axis – Weight**:
  + Represents the **proportion of total capital** allocated to each stock.
  + The total always sums to **1 (100%)**, enforcing the full investment constraint.
* **Colored Areas**:
  + Each color corresponds to a specific stock (identified in the legend).
  + The **thickness** of each colored band shows the weight of that stock in the portfolio at a given time.

**Interpretation:**

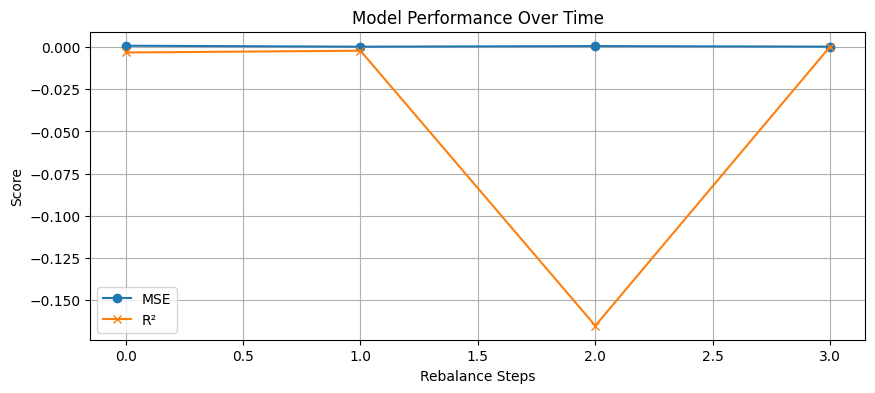
* The chart highlights a **dynamic allocation strategy**. Stock weights change significantly at each rebalance, showing that:
  + The model adjusts its preferences based on new predictions.
  + **No single stock consistently dominates**, due to the 20% max weight constraint.
* Some stocks appear intermittently, suggesting they were only selected during periods when their **predicted returns were strong** relative to their risk.
* The visual also reflects **diversification** — multiple stocks are typically included at each period, reducing reliance on any single asset.

**Strategic Insight:**

* The portfolio does **not remain static**; rather, it adapts based on the model's forecast and historical volatility.
* This aligns with an **active investment strategy**, where capital is reallocated to potentially outperforming assets each year.
* The weight capping mechanism ensures **risk control** and avoids overexposure to any one stock.

**Conclusion:**

This visualization effectively communicates how the machine learning model’s predictions and risk-aware optimization **drive real-world investment decisions**. The flexible and data-driven portfolio construction helps achieve both **return potential** and **risk diversification**.



This line chart tracks the **predictive performance** of the neural network model across **each portfolio rebalancing step** (from 2020 to 2024). It evaluates how well the model forecasted next-day stock returns using two standard regression metrics:

**Key Elements:**

* **X-axis – Rebalance Steps**:
  + Represents each annual rebalancing period when the model was retrained with updated data.
* **Y-axis – Score**:
  + Measures the quality of model predictions using:
    - **MSE (Mean Squared Error)** – Lower is better. Measures average squared prediction error.
    - **R² (R-squared)** – Closer to 1 is better. Indicates how well the model explains variance in returns.
* **Blue Line (MSE)**:
  + Stays very close to **zero**, suggesting relatively small average prediction errors across periods.
* **Orange Line (R²)**:
  + Shows **fluctuations** in performance, including one period (step 2) where R² is **negative**, indicating the model performed worse than simply predicting the average return.

**Interpretation:**

* The model generally achieves **low MSE**, suggesting it avoids large prediction errors.
* However, **R² scores vary**, highlighting that while errors are small, the model's ability to explain return variation is **inconsistent** over time.
  + A **negative R²** reflects a poor fit during that period, possibly due to market volatility, regime shifts, or overfitting.

**Strategic Insight:**

* This mixed performance reflects the **challenges of return prediction**, where market noise and external shocks can weaken model accuracy.
* Despite some weak periods, the portfolio still achieved strong overall returns, suggesting that **even moderately predictive models** can contribute to profitable strategies when combined with effective **risk management** and **optimization**.

**Conclusion:**

This chart underscores the **importance of model monitoring** in financial ML applications. It reinforces the need for **robust validation**, regular retraining, and **complementary risk controls** to ensure strategy resilience—even when predictive performance varies.

