Exploring Unsupervised Learning for Stock Returns Forecasting

Ismail Berbache, Alessandro Morosini, Ellie Yang,
Michelle Zhou, Steve Zhou

MIT MANAGEMENT SLOAN SCHOOL

Context

Problem Overview

Data

- **Source:** Yahoo Finance historical data for FTSE 250 index stocks
- **Coverage**: 10 years of trading data
- **Variables**: Open, High, Low, Close, Adjusted Close, Volume
- **Focus**: Adjusted Close prices (accounts for dividends, splits, etc.)

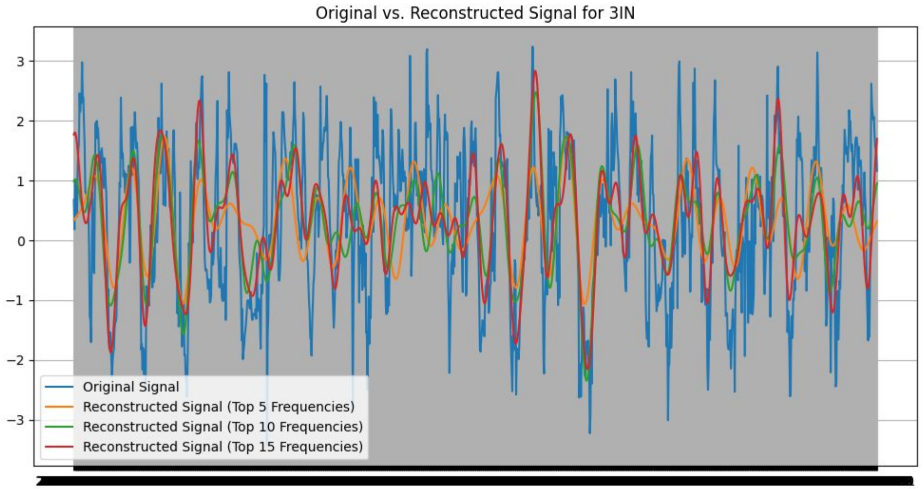
Preprocessing

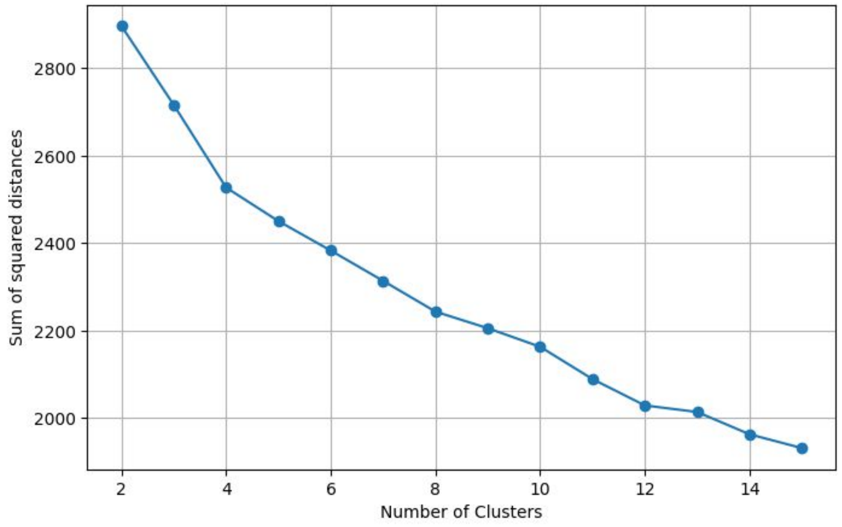
- **Removed** tickets with more than 20% missing values
- **Interpolated** moderate missing segments using time-based methods
- **Fixed outliers** and anomalies by inspection
- **Standardized** time series via rolling mean/standard deviation

Clustering Time-Series

Clustering with Fourier Transform

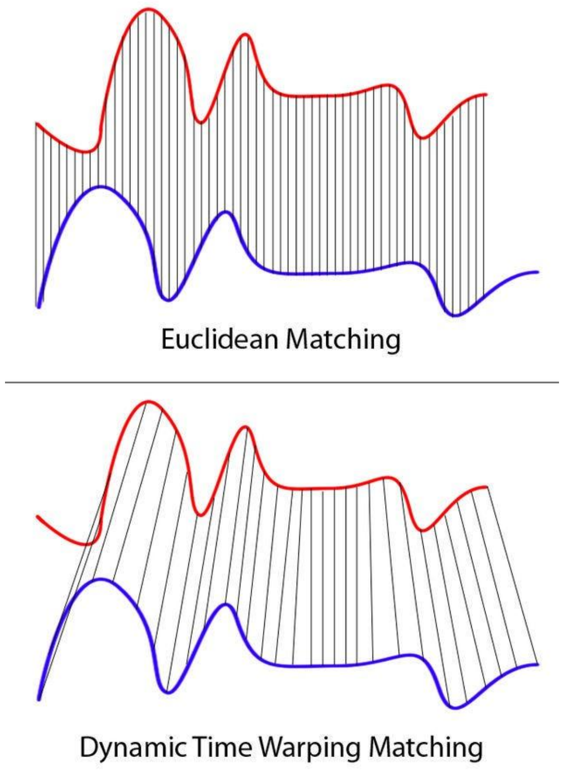
- Fast Fourier Transformation (FFT) extracts frequency signals
  - Choice of top-n subsets: tradeoff between information loss and dimensionality
  - We use top 10 most important frequencies, ranked by the norm of the complex magnitudes
- We attempted two clustering methods on the 20 coefficients:
  - K Means clustering: scree plot to use 8 clusters
  - Hierarchical clustering: group until 8 clusters remain
- Problem of using Euclidean distance for grouping Fourier coefficients: features are NOT aligned
  - Alternative way to compute set difference: a magnitude-weighted sum of frequency differences for the closest match



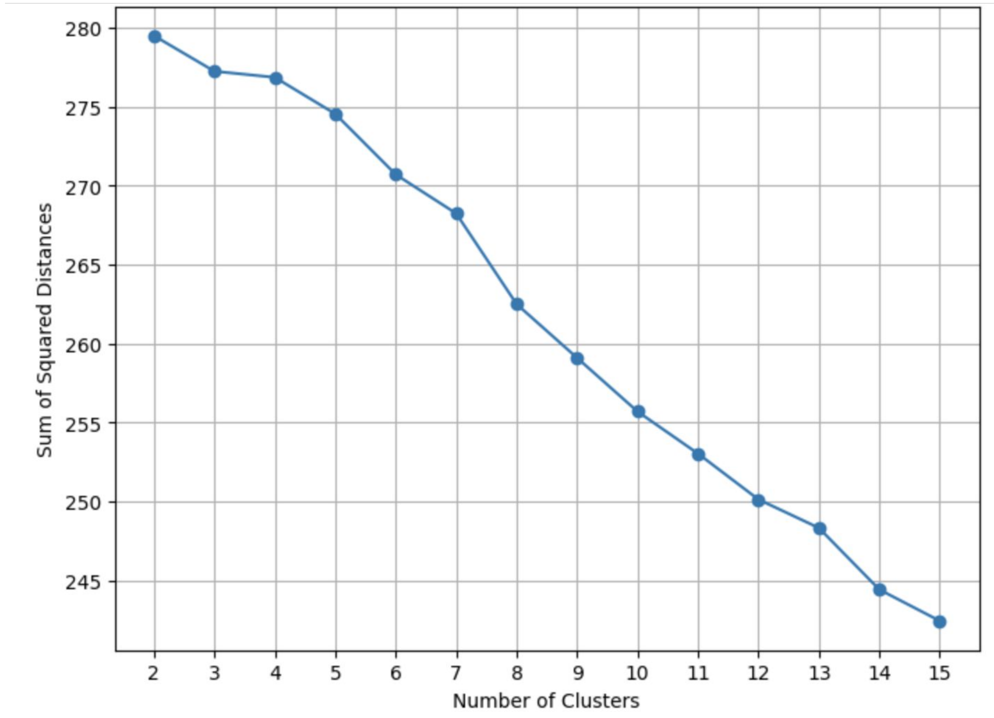


Dynamic Time Wrapping (DTW)

- Measures similarity between time-series sequences, even if they vary in speed or timing.
- Aligns sequences by stretching or compressing time to minimize the distance between them.
- Robust for comparing sequences with different lengths or misaligned patterns.
- Uses dynamic programming to compute the optimal alignment path efficiently.
- Computationally heavy!



DTW K-Means Clustering



- We only run this clustering on our training set, i.e data points from 2017 to 2021
- Linear shape of scree plot! Intuitively this could be due to the dimensionality of time-series space
- Elbow method doesn't work here. We choose k = 8 for biggest marginal decrease + interpretability

Predictive Models

Methodology Overview

Train-Test Split:

- Training Period: January 2017 – December 2021
- Testing Period: January 2022 – December 2022

Experiment setup:

- Individual Data: Train and evaluate using only the historical data of each stock
- Global Data: Combine all stocks into a single dataset for training and test on individual stocks
- Clustered Data: Group stocks into clusters. Train on cluster data and test individually

Evaluation Metrics:

- Root Mean Square Error (RMSE): Measures the average magnitude of prediction errors
- Mean Absolute Percentage Error (MAPE): Quantifies prediction accuracy as a percentage of actual values

ARIMA (baseline)

- ARIMA (AutoRegressive Integrated Moving Average), a **statistical** method for time series
- Designed for **univariate analysis**
- Only train the model **on individual stocks as baseline**, not on global and clusters.

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_q \varepsilon_{t-q}$$

Baseline ARIMA model performance:

RMSE 191.46 and MAPE 24.85%

MIT MANAGEMENT SLOAN SCHOOL

LightGBM

Feature Engineering:

- Over different window size: 5, 10 and 20
- Lagged value of stock prices, moving averages, rolling standard deviations, max and min
- Rate of Change (ROC), Relative Strength Index (RSI)

Hyperparameter Grid Search:

- 'feature_fraction': 0.9, 'learning_rate': 0.05, 'n_estimators': 300, 'num_leaves': 15, 'reg_alpha': 0.1, 'reg_lambda': 0.1

LightGBM Results

- Models were trained on individual stocks, all stocks (global), and stocks grouped by clusters
- Applied **6** different clustering methods to group stocks
- Best performance achieved with **global** data; **DTW with 5 clusters** ranked second

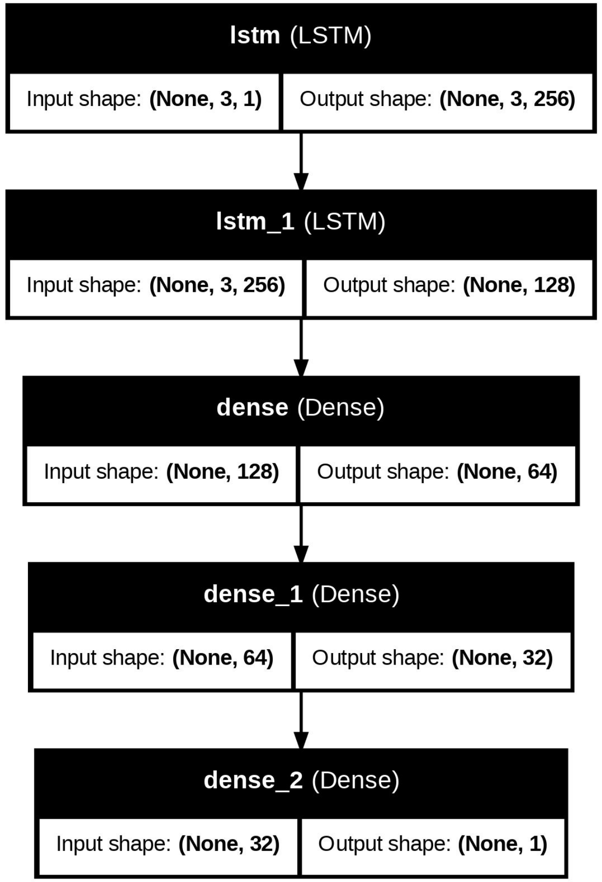
	By Individual	By Global	By Cluster (DTW)		By Cluster (FFT)			
Metric	**Individual**	**Global**	**DTW_5**	**DTW_8**	**kmean_5**	**kmean_8**	**hier_5**	**hier_8**
RMSE	34.39	17.23	17.89	21.86	19.82	20.79	18.42	18.49
MAPE	3.99%	1.76%	1.79%	1.97%	1.87%	1.97%	1.83%	1.85%

Table 1: LightGBM Model performance comparison

MIT MANAGEMENT SLOAN SCHOOL

LSTM (Long Short-Term Memory networks)

- **Data preparation:** convert data into windowed format (window size = 3)
- **Model Architecture:**
  - Two stacked LSTM layers (256 and 128 units)
  - Two dense layers (64 and 32 units, ReLU activation)
  - Output layer: Single neuron for regression)
  - (The parameters are tuned when training on individual stock)
- Experiment with 5 and 8 Clusters: 5 clusters perform better



MIT MANAGEMENT SLOAN SCHOOL

Result Analysis

(**Baseline ARIMA:** RMSE 191.46 and MAPE 24.85%)

Training Configuration	Average RMSE	Average MAPE (%)
Individual Stock Training	34.39	3.99
Clustered Stock Training	17.89	1.79
All Stock Training	17.23	1.76

Table 1: LightGBM Performance

Training Configuration	Average RMSE	Average MAPE (%)
Individual Stock Training	49.08	3.80
Clustered Stock Training	45.45	3.25
All Stock Training	27.33	2.57

Table 2: LSTM Performance

Future Work

Expand Dataset

Include more stocks for a deeper evaluation of clustering advantages in larger and noisier datasets

More models and Features

Experiment with a wider variety of models and features. Investigate if some architecture or higher dimensional data would prefer clustering data as training data.

Develop Trading Strategy

Design and test trading strategies based on the model predictions, and evaluate on financial returns to evaluate the practical utility of the models in real-world applications.

THANK YOU

MIT MANAGEMENT SLOAN SCHOOL