Data Quality Report

The data set that is going to be used in this paper would be the stock price and volume data in the historical measurement combined with a variety of technical indicators that are computed using rolling window. Preliminary data quality analysis was performed to determine the completeness, consistency and time coverage before exploratory data analysis and data modelling. Comprehensively, the dataset is highly oriented in integrity and it can be downstreamed to analytical activities.

The gaps in values were noticed mostly in the engineered features including moving averages, momentum indicators, the RSI, MACD, and volatility measures. These omissions were fairly localized, and they typically formed less than three percent of the data per feature. This was structural not random, the computer had to initialize the rolling window computations at the start of each time series to achieve the missingness. This meant that there was no imputation of core financial variables. There are a few rows in the data that lacked enough historical background so they were either avoided during preprocessing or they were simply not needed during the model training, allowing the calculation of features to be valid.

A check on record distinctiveness affirmed that there were no duplicate records. The observation is based on unique combination of stock identifier and trading date and thus it maintains integrity of time series analysis. Temporal coverage of all instruments runs between early 2021 to the end of 2023, with daily frequency and dates matching between price, volume and derived indicators. This consistency makes it possible to compare across time and assets.There were no obvious data abnormalities, contradictions, or corruption. The small problems with the indicator initiation were dealt with using the usual preprocessing methods without losing data volume and representativeness. The data is purged, properly formatted, and can be used as a credible base of the exploratory analysis and predictive modeling.