**IT3191E MACHINE LEARNING AND DATA MINING**

**PROJECT PROPOSAL**

**Class: 139399**

**Lecturer: Ph.D. Nguyen Nhat Quang**

|  |  |  |
| --- | --- | --- |
| **Student Name** | **Student ID** | **Email** |
| Nguyen Khanh Trung | 20205133 | [trung.nk205133@sis.hust.edu.vn](mailto:trung.nk205133@sis.hust.edu.vn) |
| Nguyen Kim Tuyen | 20205196 | [tuyen.nk205191@sis.hust.edu.vn](mailto:tuyen.nk205191@sis.hust.edu.vn) |

***Title***

Pair selection for statistical arbitrage by pair trading using clustering

***Description***

Pair trading is a popular trading strategy used in financial markets. The logic behind pairs trading is to trade pairs of stocks which are deemed **cointegrated**.

Cointegration is used to identify stocks that have a long-term relationship where their prices **move in tandem over time**, even though they may **diverge** in the short-term.

When two stocks are cointegrated, they **tend to revert to** their long-term relationship after any short-term divergence. This provides an opportunity for a pair trader to take advantage of any short-term price discrepancies between the two stocks.

In this project, we investigate the application of clustering algorithms to find appropriate pairing in the stock market for pair trading strategy. Our aim for the project is to demonstrate the potential of data mining techniques for improving the performance of pair trading strategies and provides a valuable contribution to the field of financial engineering.

***Input***

* Daily prices of the stock universe constrained to the HNX and HOSE
* Domain prior knowledge: industry, financial health, market cap…

***Output***

List of eligible pairs for the pair trading strategy. (With visualization)

***Approach***

* Kmeans vs DBSCAN clustering algorithms
* ADF test for cointegration discoveries

***Dataset***

* Daily prices dataset of stock universe from HNX, HOSE
* Industry dataset crawled from <https://vietstock.vn/>