

Dart Quant Strategy Series

End-to-End

Pairs Trading Strategy

Choi Munseok, 12th

Choi Wonjun, 13th

May

2024

목차

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

21

1. 소개
2. 사전연구 검토
3. 방법론과 데이터
   1. 주성분분석
   2. 군집화 알고리즘
   3. 성능 평가
   4. 칼만필터(Kalman Filter)
   5. 정상성 검정
   6. 다중 페어트레이딩
   7. 데이터 설명
4. 실증연구
   1. 분석기간
   2. 군집화 기반 앙상블 페어선택 알고리즘
   3. 칼만필터를 활용한 스프레드 동적 추정 알고리즘
5. 결론
6. 부록

**DART 2024 First Semester Pairs Trading Project**

**Choi Munseok**

**cmschs0301@knu.ac.kr**

**Choi Wonjun**

**song1002@knu.ac.kr**

**[Github]**

**https://github.com/DART-KNU/Pairs-Trading-from-Beginning-to-End**

**Cover image source : Space Center Houston**

1. **Introduction**

오늘날

1. **Literature Review**

**III. Methodology and Data**

**3.1 Principal Component An**

**IV. 실증연구**

**4.1 분석기간**

Train/Validation/Test 데이터셋으로 6:2:2 비율로 나눠 연구를 진행하였다.

**4.2 군집화 기반 앙상블 페어선택 알고리즘**

일반적으로 페어트레이딩에서 페어선택 방법은 모든 유가증권 간의 조합을 고려하여 모든 가능한 후보 페어를 선택하는 것이다. 사용가능한 유가증권이 n개라고 했을 때, 가능한 조합의 수는 이다. 이때, 두가지 문제가 발생하게 된다. 첫째, 모든 가능한 조합에 대한 평균회귀 테스트의 시간복잡도는 이다. 따라서 고려되는 유가증권이 늘어날수록, 계산량이 급격하게 증가하게 된다. 둘째, 다중 검정 문제가 발생한다. 하나의 테스트에서 귀무가설을 기각할 확률, 즉 제 1종 오류를 저지를 확률은 이다. 하나의 테스트에서 귀무가설을 기각하지 않을 확률은 이며, 모든 m개의 테스트에서 귀무가설을 기각하지 않을 확률은 이다. 적어도 하나의 테스트에서 귀무가설을 기각할 확률은 가 된다. 즉, 각각의 가설에 대해 유의수준(1종오류)인 검정을 동시에 수행할 경우 전체 오류율이 매우 커지게 된다. 해당부분은 Family-wise type 1 error rate(FWER)로 잘 알려져 있다. 다중 비교 문제의 영향을 완화하는 방법은 두가지이다. 첫째, 본페로니 교정(Bonferroni correction)과 같은 다중 보정테스트를 적용하는 방법이다. 본페로니 교정은 유의수준을 총 검정의 수로 나누어서 개별 검정을 보정하는 방법이다. 하지만, Harlacher[8]는 본페로니 보정이 페어선택에 있어 너무 보수적이 되어 실제로 페어의 발견을 방해한다는 것을 밝혔다. 둘째, 통계적 테스트의 수를 줄이는 것이다. 본 장에서는 Simão Moraes Sarmento[9]가 제안한 군집화 기반 앙상블 페어선택 알고리즘을 사용하여 통계적 테스트의 수를 줄인다. 또한, 해당 알고리즘에 대한 저자의 개인적인 생각을 바탕으로 수정한 새로운 방법론과 성능을 비교한다. 군집화 기반 앙상블 페어선택 알고리즘은 (1) 비슷한 위험요소에 노출된 유가증권끼리 묶여 함께 움직일 가능성이 높으며 (2) 필요한 통계적 테스트의 수가 급감하는 장점이 있다.

**4.2.1 알고리즘1**

Simão Moraes Sarmento[9]가 제안한 알고리즘은 비지도 학습 알고리즘을 사용하여 의미 있는 자산 군집을 추론하여 쌍을 선택하는 것이다. 순서는 다음과 같다.

1. 차원축소 2) 군집화 알고리즘 적용 3) 페어 선택

**4.2.2 알고리즘2**

Simão Moraes Sarmento[9] 제안한 군집화 기반 앙상블 페어선택 알고리즘은 설정된 기간 내에서 nxm(n은 주식 수, m은 기간)형태의 행렬 A을 설정한다. 행렬 A의 공분산행렬은 mxm 행렬이 되고, 공분산 행렬에 주성분 분석을 적용해 m차원의 m개 고유 벡터 집합을 생성한다. 여기서 고윳값에 따른 고유 벡터의 수 k를 설정하고, k개의 고유벡터의 집합으로 구성된 mxk행렬을 행렬 A에 행렬곱 해준다. 이렇게 나온 nxk 행렬에 대해 군집화를 진행한다.

**4.2.3 실증분석**

앞서 소개한 1번 알고리즘과 2번 알고리즘을 상관계수를 제곱한 값 즉 결정계수를 사용하여 평가한다. 상관계수 자체는 두 변수 간의 선형 관계의 강도와 방향을 나타낸다. 시계열 데이터의 군집화 관점에서 결정계수를 사용하여 군집 내 종목들이 얼마나 유사한지를 평가한다. Train 데이터 셋에 알고리즘을 적용하여 군집화를 마친 후 Validation 데이터셋에서 알고리즘의 성능을 평가한다. 알고리즘의 성능을 5회 측정하여 평균한 값은 <표5>와 같다.



<표 5> 군집화 기반 앙상블 페어선택 알고리즘 성능비교

1번 알고리즘에서 DBSCAN을 사용했을 때 성능이 가장 좋게 나온 것을 확인할 수 있다. 하지만 이때, 군집의 수가 1개이고 군집내 종목의 수가 2이므로 본 연구의 목적에 부합하지 않는다. 따라서 두번째로 성능이 좋으며 군집의 개수와 군집내 종목의 개수가 어느정도 보장되는, 2번 알고리즘의 Hierarchical을 사용하여 분석을 진행한다. 군집내 구성종목은 <표6>과 같다.

<표 6> 군집내 구성종목

Validation Set으로 군집내 페어가 선정되는지 또한 확인하였다. 군집3과 군집5를 제외한 모든 군집에서 many-to-many 페어가 선정이 되었으므로, 군집 5를 제외한 나머지 군집으로 거래를 진행한다.

**4.3 칼만필터를 활용한 스프레드 동적 추정 알고리즘**

**4.3.1 one-to-one 페어 선정**

우선 군집내 자산들 간 Engle-Granger 두 단계 공적분 검정(Engle-Granger Two-Step Cointegration Test)를 수행하여 자산간의 공적분 관계를 식별하고 이를 통해 군집내 one-to-one 페어를 선정합니다. 검정 순서는 다음과 같습니다. 1) OLS 회귀분석: 두 자산의 시계열 데이터를 사용하여 단순 회귀분석(OLS: Ordinary Least Squares)를 수행합니다. 2) 잔차 계산: 회귀모델의 잔차(residuals)를 계산합니다. 3) ADF 테스트 수행: 잔차에 대해 ADF(Augmented Dickey-Fuller) 검정을 사용하여 잔차가 정상성을 갖는지 확인합니다.

* + 1. **서브군집 분할 알고리즘**

다음은 군집내 자산들을 두 그룹으로 분할하기 위한 알고리즘을 설명합니다. <그림1>에서 설명하는 것 처럼 공적분 관계를 가지는 one-to-one 페어들을 대상으로 서브군집 분할 알고리즘을 사용하여 자산 쌍을 두 그룹으로 나누는 방식을 사용합니다. 우선, 군집내 각 주식을 노드로, 공적분관계를 갖는 노드들을 간선으로 잇는 그래프를 생성합니다. 이는 비 연결 무방향 그래프입니다. 연결된 간선이 없는 노드들은 제거를 하고 간선이 연결된 노드들을 대상으로 서브 군집 분할 알고리즘을 수행합니다. 알고리즘의 의사코드(pseudocode)는 <코드1>과 같습니다.

도표, 라인, 원, 디자인이(가) 표시된 사진

자동 생성된 설명

<그림 1> 서브군집 분할



<코드 1> 서브군집 분할 알고리즘

위 코드는 크게 세차례의 과정을 거칩니다. 1) 초기화 2) 그룹 분할 3) 결과 출력 및 반환

1. 초기화

가) 첫번째 자산 쌍을 이용하여 초기 설정을 합니다.

나) 첫 번째 자산을 ‘group1’에 추가하고, 두 번째 자산을 ‘group2’에 추가합니다.

1. 그룹 분할:
2. 자산 쌍의 첫 번째 자산이 이미 ‘group1’에 있고, 두 번째 자산이 아직 어떤 그룹에도 속하지 않는다면 이를 ‘group2’에 추가합니다.
3. 반대로, 첫 번째 자산이 ‘group2’에 있고, 두 번째 자산이 아직 어떤 그룹에도 속하지 않으면 이를 ‘group1’에 추가합니다.
4. 만약 자산 쌍의 두 자산이 어느 그룹에도 속하지 않는다면, 각각 다른 그룹에 추가합니다.
5. 결과 출력 및 반환

**4.3.3 칼만필터를 활용한 스프레드 동적 추정 알고리즘**

본 알고리즘은 칼만 필터를 활용하여 공적분 관계를 가지는 자산 쌍의 스프레드를 동적으로 추정하는 과정입니다. 자산 쌍의 가격 데이터를 로그 변환하여 독립 변수와 종속 변수를 설정합니다. 독립 변수로 관측 행렬을 생성하고, 초기 상태 평균과 공분산, 상태 전이 행렬, 관측 행렬 등을 설정하여 칼만 필터를 초기화합니다. 여기서 델타()는 1e-5로 설정되며, 상태 공분산은 로 계산됩니다. 필터를 적용하여 동적으로 변하는 회귀 계수를 추정하고, 이 계수를 이용해 각 시점에서의 스프레드를 계산합니다.

**4.3.4 Many-to-Many 페어 스프레드 추정**

one-to-one 페어의 스프레드는 칼만필터를 활용한 동적 추정 알고리즘을 활용해 결정한다. one-to-one 페어들을 사용하여 4.3.2의 서브 군집 분할 알고리즘을 활용한 서브 군집에 따라 many-to-many 페어를 결정한다. Many-to-many 페어를 구성하는 각 one-to-one 페어의 비중은 다음과 같이 결정된다. 페어트레이딩에서는 스프레드의 평균회귀를 기대하며 베팅하기 때문에 페어의 공적분 관계가 유지되는 것이 중요하다. 베팅 사이즈를 결정하기 위해, 각 페어 Engle-Granger 두 단계 공적분 검정의 ADF-test에서 결정된 p-value를 활용한다.

1. 먼저, 각 p-value의 역수에 로그를 씌운 값을 계산합니다:
2. 모든 항목에 대해 위 값을 합산합니다:
3. 각 항목의 비중은 다음과 같이 계산됩니다:

각 one-to-one페어에 앞서구한 weight를 적용시켜 만든 many-to-many 페어에 대해 ADF-test를 진행하고, 정상성이 있다고 판단되면 해당 many-to-many 페어에 대해 거래를 진행한다.

**4.3.5 거래 전략**

1) 최대 보유기간 설정: 반감기

1. 손절 기준: 3표준편차
2. 매수 시그널: 2 표준편차 이탈 후 진입시
3. 매도 시그널: 평균 (2 표준편차 x 0.1)

1)에서 반감기는 다음과 같이 계산된다.

AR(1) 모델임을 나타냅니다.

여기서 는 현재 시점의 값, 는 이전 시점의 값, 는 자기회귀 계수, 는 백색 잡음(white noise)을 의미합니다. AR(1) 모델을 데이터에 적합한 후 자기회귀 계수를 추정합니다. 반감기는 값이 절반으로 감소하는 데 걸리는 시간을 의미하며, 지수 감소(exponential decay) 모델을 기반으로 계산할 수 있습니다. AR(1) 모델의 반감기는 다음과 같은 수식을 통해 구할 수 있습니다.

반감기 는 다음과 같이 계산됩니다:

* 는 자기회귀 계수입니다.

1. 자기회귀 계수 를 사용하여 값이 절반으로 감소하는 시점을 구합니다.

2. 자연 로그를 사용하여 지수 감소의 특성을 반영합니다.

3. 는 값이 절반으로 감소하는 지점을 나타냅니다.

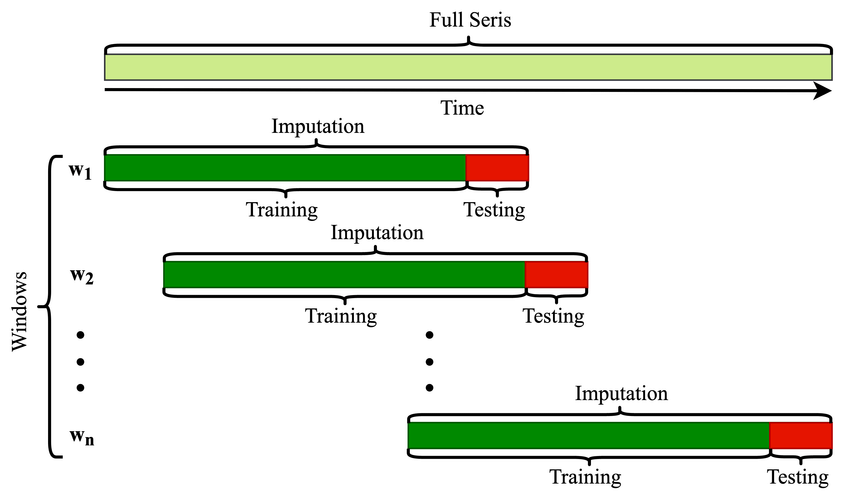
**4.3.5 백테스트**

1) 벤치마크 설정

페어 트레이딩 전략은 시장 중립 전략으로, 전략의 성과 평가 시 벤치마크로서 무위험 수익률을 설정하는 대신, 절대 기준 수익률을 0으로 설정합니다. 이는 시장의 상승 또는 하락과 관계없이 전략의 순수한 알파(α)를 측정하고, 시장의 방향성에 대한 노출을 최소화한 상태에서의 성과를 평가하는 데 중점을 둡니다.

1. 롤링 윈도우 분석(Rolling Window Analysis)

롤링 윈도우 분석은(Rolling Window Analysis) 슬라이딩 윈도우sliding window 방식을 사용하여 전략의 성과를 평가하는 방법입니다. <그림n>과 같이 일정 크기의 시간 구간(윈도우)를 설정하고, 이 윈도우를 일정 간격으로 이동시키면서 각 구간마다 전략의 성과를 평가합니다. 본 연구에서 사용된 첫번째 window의 train data는 1999.05.14에서 2016.01.04이고 test data는 2016.01.05에서 2020.03.17이다. 두번째 window부터 256영업일만큼 일정 간격으로 이동하면서 각 구간마다 전략의 성과를 평가합니다.



<그림n> Rolling Window Analysis

1. 벡테스트 결과

모든 Window에서 벤치마크를 상회하는 모습을 볼 수 있음

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Window** | **1** | **2** | **3** | **4** | **5** |
| **구간 수익률(1030영업일: 약 4년, 소수 3째 자리에서 반올림)** | **0.342%** | **1.36%** | **5.06%** | **1.37%** | **1.77%** |

**V. 결론**

**VI. 부록**

**5.1. 알고리즘1 군집화 결과**

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Cluster Method** | **Cluster** | **Stocks** |
| **algorithm1 - window1** | **KMeans** | **1** | **LG, HD한국조선해양, 삼성중공업, 현대제철, HD현대미포, 동국홀딩스, 한진중공업홀딩스** |
| **KMeans** | **2** | **삼성전자, SK하이닉스, 삼성SDI, 삼성전기** |
| **KMeans** | **3** | **동원산업, 사조대림, 사조산업, 신라교역, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **4** | **삼성화재, 기업은행, 코리안리** |
| **KMeans** | **5** | **현대건설, 삼부토건, 코오롱글로벌, 계룡건설, 동부건설, HL D&I** |
| **KMeans** | **6** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **KMeans** | **7** | **현대차, 기아, 현대모비스, 한국앤컴퍼니** |
| **KMeans** | **8** | **동서, 콜마홀딩스, 태경케미컬, 무학, KH 필룩스, 한국내화, 황금에스티, 인팩, 부국철강** |
| **KMeans** | **9** | **CJ, 에스원, 제일기획, 신세계, 오리온홀딩스, 빙그레, SBS** |
| **KMeans** | **10** | **태광산업, 고려제강, BYC, 대한화섬, 비비안** |
| **KMeans** | **11** | **SK가스, E1, 삼천리, 한국쉘석유, 예스코홀딩스, 경동인베스트** |
| **KMeans** | **12** | **유한양행, 한미사이언스, 대웅, 녹십자홀딩스, 동아쏘시오홀딩스, 환인제약** |
| **KMeans** | **13** | **POSCO홀딩스, 한국전력** |
| **KMeans** | **14** | **고려아연, S-Oil, LS, 롯데케미칼, 금호석유, KCC, OCI홀딩스** |
| **KMeans** | **15** | **SK텔레콤, KT** |
| **KMeans** | **16** | **쌍용C&E, 한일현대시멘트, 동양, 성신양회, 진흥기업, 까뮤이앤씨, 남광토건, 범양건영** |
| **KMeans** | **17** | **GS건설, DL, HDC** |
| **KMeans** | **18** | **한화솔루션, SKC, 한화, 효성, 롯데정밀화학, SK디스커버리, 풍산홀딩스, 코오롱** |
| **KMeans** | **19** | **에스엘, 화신, 대원강업, 서연, 에스제이지세종, 디와이** |
| **KMeans** | **20** | **신영증권, 부국증권, 유화증권, 상상인증권** |
| **KMeans** | **21** | **보령, JW중외제약, 부광약품, 광동제약, 종근당홀딩스, 삼진제약, 일양약품, 일성아이에스, 동화약품** |
| **KMeans** | **22** | **한온시스템, 한국단자, 삼영전자, 대덕, 한국전자홀딩스** |
| **KMeans** | **23** | **DB손해보험, 현대해상, 롯데손해보험, 한화손해보험, 제주은행, 흥국화재** |
| **KMeans** | **24** | **만호제강, 성창기업지주, 방림, SG글로벌, 대한방직, SUN&L** |
| **DBSCAN** | **1** | **SK텔레콤, KT** |
| **Hierarchical** | **1** | **삼성전자, SK하이닉스, POSCO홀딩스, 삼성SDI, 한국전력, 삼성전기, SK텔레콤, KT** |
| **Hierarchical** | **2** | **S-Oil, 영풍, 태광산업, 고려제강, 송원산업, 경방, KCTC, 크라운해태홀딩스, 신영와코루, 카프로** |
| **Hierarchical** | **3** | **BYC, 만호제강, 대한화섬, 성창기업지주, 방림, 대한방직, SUN&L, 비비안** |
| **Hierarchical** | **4** | **현대차, 기아, 현대모비스, 한온시스템, 한국앤컴퍼니** |
| **Hierarchical** | **5** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **Hierarchical** | **6** | **동서, 콜마홀딩스, 태경케미컬, 에스제이지세종, KH 필룩스, 한국내화, 황금에스티, 인팩, 부국철강** |
| **Hierarchical** | **7** | **삼성화재, 기업은행, DB손해보험, 현대해상, 코리안리, 롯데손해보험, 한화손해보험, 제주은행, 흥국화재** |
| **Hierarchical** | **8** | **동원산업, 사조대림, 사조산업, 신라교역, 사조오양, 한성기업, 동원수산** |
| **algorithm1 - window2** | **KMeans** | **1** | **KG모빌리티, 세아베스틸지주, 제주은행, LS네트웍스, GS글로벌, 신원, 한신공영, 일성건설** |
| **KMeans** | **2** | **CJ, 제일기획, 오뚜기, 오리온홀딩스, 대상, 빙그레, 삼양홀딩스, 한섬, SBS** |
| **KMeans** | **3** | **미래에셋증권, NH투자증권, 삼성증권, 대신증권** |
| **KMeans** | **4** | **현대차, 기아, 현대모비스** |
| **KMeans** | **5** | **에스엘, 화신, 서연, 에스제이지세종, 디와이** |
| **KMeans** | **6** | **아모레G, 롯데지주, 농심, 롯데칠성, 영풍, 퍼시스, 한일홀딩스, 신도리코, 남양유업** |
| **KMeans** | **7** | **동원산업, 사조대림, 사조산업, 신라교역, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **8** | **세아제강지주, 극동유화, 한국특강, 한일철강, 부국철강** |
| **KMeans** | **9** | **SK하이닉스, 카카오, 동원시스템즈, 디아이, 신성이엔지, 와이투솔루션, 콤텍시스템, 미래산업** |
| **KMeans** | **10** | **DN오토모티브, 태광산업, 고려제강, 아세아, KISCO홀딩스, CR홀딩스, 대한제분, KPX케미칼, 하이트진로홀딩스, 경방** |
| **KMeans** | **11** | **대한항공, LS ELECTRIC, 호텔신라, 롯데정밀화학, 아시아나항공, SK디스커버리, 한진, 코오롱, 한솔홀딩스** |
| **KMeans** | **12** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **KMeans** | **13** | **쌍용C&E, 한일현대시멘트, 성신양회** |
| **KMeans** | **14** | **삼성E&A, 현대건설, GS건설, DL, HDC** |
| **KMeans** | **15** | **SK가스, E1, 삼천리, 한국쉘석유, 예스코홀딩스, 신대양제지, 경농, 태경산업, 경동인베스트, 미창석유** |
| **KMeans** | **16** | **삼성화재, 기업은행, 코리안리** |
| **KMeans** | **17** | **DB손해보험, 현대해상, 롯데손해보험, 한화손해보험, 흥국화재** |
| **KMeans** | **18** | **DI동일, BYC, 만호제강, 대한화섬, 성창기업지주, 방림, 대한방직** |
| **KMeans** | **19** | **삼성전자, 삼성SDI, 삼성전기, 한화에어로스페이스, 대덕** |
| **KMeans** | **20** | **한국전력, KT&G, SK텔레콤, KT** |
| **KMeans** | **21** | **SNT다이내믹스, 에이프로젠, 현대비앤지스틸, SIMPAC, STX, 동원금속, 대양금속, 케이비아이동국실업** |
| **KMeans** | **22** | **휴스틸, NI스틸, 동양철관, 문배철강** |
| **KMeans** | **23** | **삼부토건, 동양, 코오롱글로벌, 진흥기업, 계룡건설, 동부건설, HL D&I, 까뮤이앤씨, 남광토건, 범양건영** |
| **KMeans** | **24** | **유한양행, 한미사이언스, 녹십자, 대웅, 녹십자홀딩스, 동아쏘시오홀딩스, 종근당홀딩스** |
| **KMeans** | **25** | **한온시스템, 한국가스공사, 에스원, 신세계** |
| **KMeans** | **26** | **POSCO홀딩스, HD한국조선해양, S-Oil, 현대제철, 동국홀딩스** |
| **KMeans** | **27** | **영진약품, 진원생명과학, 삼성제약, 에이프로젠바이오로직스, CJ씨푸드, 팜젠사이언스, 오리엔트바이오** |
| **KMeans** | **28** | **한국앤컴퍼니, 넥센타이어, 한국단자, 대원강업, 모토닉, 한국무브넥스, 인지컨트롤스, 화승코퍼레이션, 인팩, 유성기업** |
| **DBSCAN** | **1** | **S-Oil, 동원산업, 사조대림, 사조산업, 신라교역, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **1** | **삼성전자, 현대차, 기아, 삼성SDI, 현대모비스, 삼성전기** |
| **Hierarchical** | **2** | **에스엘, 화신, 대원강업, 모토닉, 서연, 에스제이지세종, 한국무브넥스, 인지컨트롤스, 화승코퍼레이션, 유성기업** |
| **Hierarchical** | **3** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **Hierarchical** | **4** | **삼성화재, 기업은행, DB손해보험, 현대해상, 코리안리, 롯데손해보험, 한화손해보험, 흥국화재** |
| **Hierarchical** | **5** | **한국전력, KT&G, SK텔레콤, KT, 한국가스공사** |
| **Hierarchical** | **6** | **HLB글로벌, 휴스틸, NI스틸, 광명전기, 동양철관, 대양금속, 문배철강** |
| **Hierarchical** | **7** | **동원산업, 사조대림, 사조산업, 신라교역, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **8** | **DI동일, BYC, 만호제강, 대한화섬, 성창기업지주, 방림, 대한방직** |
| **algorithm1 - window3** | **KMeans** | **1** | **POSCO홀딩스, 두산에너빌리티, 한화오션, HD한국조선해양, 삼성중공업, 현대제철, HD현대미포, HD현대인프라코어, 동국홀딩스** |
| **KMeans** | **2** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **3** | **일성아이에스, 삼일제약, 한독, 제일파마홀딩스, 일동홀딩스** |
| **KMeans** | **4** | **코오롱글로벌, 계룡건설, 동부건설, 화성산업, HL D&I** |
| **KMeans** | **5** | **NI스틸, 동양철관, 문배철강** |
| **KMeans** | **6** | **KCC, 현대엘리베이, SK네트웍스, 대한유화, 세아제강지주, 세아베스틸지주, 풍산홀딩스** |
| **KMeans** | **7** | **한국전력, KT&G, SK텔레콤, KT, 한국가스공사** |
| **KMeans** | **8** | **삼화콘덴서, 삼화전기, 삼화전자** |
| **KMeans** | **9** | **DB하이텍, 송원산업, 현대비앤지스틸, 이수화학, AK홀딩스, 대창, 유니온, 서원, 이건산업, 카프로** |
| **KMeans** | **10** | **현대차, 기아, 현대모비스** |
| **KMeans** | **11** | **카카오, 엔씨소프트, 코웨이, LG유플러스, 영원무역홀딩스, 현대지에프홀딩스, 제주은행, 흥국화재, HS애드, 대구백화점** |
| **KMeans** | **12** | **삼부토건, 동양, 진흥기업, 한신공영, 일성건설, 까뮤이앤씨, 남광토건, 범양건영** |
| **KMeans** | **13** | **한온시스템, 아모레G, 롯데지주, 농심, 에스원, 신세계, 롯데칠성, 신도리코, 남양유업** |
| **KMeans** | **14** | **삼성전자, 삼성SDI, 삼성전기, 한화에어로스페이스, 한국앤컴퍼니, 대덕** |
| **KMeans** | **15** | **SK가스, E1, 삼천리, 한국쉘석유, 예스코홀딩스, 신대양제지, 경농, 미창석유, 극동유화, 한국수출포장** |
| **KMeans** | **16** | **LG화학, LG, S-Oil, 한화솔루션, 롯데케미칼** |
| **KMeans** | **17** | **삼성화재, 기업은행, DB손해보험, 현대해상, 코리안리, 한화손해보험** |
| **KMeans** | **18** | **현대건설, 대우건설, GS건설, DL, HDC** |
| **KMeans** | **19** | **영진약품, 진원생명과학, 삼성제약, 에이프로젠바이오로직스, 팜젠사이언스, 유유제약, 오리엔트바이오** |
| **KMeans** | **20** | **신영증권, 부국증권, 유화증권, 상상인증권** |
| **KMeans** | **21** | **포스코인터내셔널, 대한전선, 대한해운, KG스틸, STX, 금호건설, 유수홀딩스, 태영건설** |
| **KMeans** | **22** | **쌍용C&E, 한일홀딩스, 한일현대시멘트, 성신양회** |
| **KMeans** | **23** | **유한양행, 한미사이언스, 녹십자, 녹십자홀딩스, 동아쏘시오홀딩스** |
| **DBSCAN** | **1** | **S-Oil, 삼화콘덴서, 삼화전기, 삼화전자** |
| **Hierarchical** | **1** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **Hierarchical** | **2** | **현대차, 기아, 현대모비스** |
| **Hierarchical** | **3** | **한국전력, KT&G, SK텔레콤, KT, 한국가스공사** |
| **Hierarchical** | **4** | **현대건설, 대우건설, GS건설, DL, HDC** |
| **Hierarchical** | **5** | **BYC, 만호제강, 대한화섬, 성창기업지주, 방림, 대한방직** |
| **Hierarchical** | **6** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **7** | **NI스틸, 광명전기, 동양철관, 문배철강, 선도전기** |
| **algorithm1 - window4** | **KMeans** | **1** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **2** | **미래에셋증권, NH투자증권, 삼성증권, 대신증권** |
| **KMeans** | **3** | **LG생활건강, CJ, 아모레G, 농심, 오뚜기, 롯데칠성, 오리온홀딩스, 대상, 동원F&B, 삼양홀딩스** |
| **KMeans** | **4** | **한일현대시멘트, 성신양회, 경농, HL D&I, 일성건설** |
| **KMeans** | **5** | **삼성전자, SK하이닉스, 삼성SDI, LG전자, 삼성전기** |
| **KMeans** | **6** | **조광피혁, 삼화페인트, 신세계 I&C, 강남제비스코, 삼양통상, 성창기업지주, 디씨엠, 방림, 까뮤이앤씨, 대한방직** |
| **KMeans** | **7** | **대원전선, 신원, 광명전기, 한국내화, 동양철관, 일신석재, 혜인, 조비, 선도전기, 인디에프** |
| **KMeans** | **8** | **코웨이, LG유플러스, 강원랜드, 에스원, 제일기획, 신세계, 빙그레, 한섬, SBS** |
| **KMeans** | **9** | **세아제강지주, 가온전선, 휴스틸, 태경산업, 황금에스티, 화성산업, 유니온, 신세계건설, 이건산업** |
| **KMeans** | **10** | **현대비앤지스틸, 대창, 서원** |
| **KMeans** | **11** | **삼화콘덴서, 삼화전기, 삼화전자** |
| **KMeans** | **12** | **삼성화재, DB손해보험, 현대해상, 코리안리** |
| **KMeans** | **13** | **POSCO홀딩스, 현대제철, 동국홀딩스** |
| **KMeans** | **14** | **한국전력, KT&G, SK텔레콤, KT, 롯데지주, 한국가스공사, 신도리코** |
| **KMeans** | **15** | **현대차, 기아, 현대모비스** |
| **KMeans** | **16** | **한온시스템, 한국앤컴퍼니, 넥센타이어, 한국단자** |
| **KMeans** | **17** | **유한양행, 한미사이언스, 녹십자, 대웅, 녹십자홀딩스, 동아쏘시오홀딩스, 종근당홀딩스, 환인제약, 한독** |
| **KMeans** | **18** | **NI스틸, 부국철강, 문배철강** |
| **KMeans** | **19** | **LG화학, LG, 고려아연, S-Oil, 한화솔루션, 롯데케미칼, OCI홀딩스** |
| **KMeans** | **20** | **HMM, 쌍용C&E, 현대엘리베이, 남해화학, 현대코퍼레이션** |
| **KMeans** | **21** | **영진약품, 일성아이에스, 삼일제약, 진원생명과학, 제일파마홀딩스, 삼성제약, 에이프로젠바이오로직스, 팜젠사이언스, 오리엔트바이오** |
| **KMeans** | **22** | **신한지주, 기업은행** |
| **DBSCAN** | **1** | **S-Oil, 동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **1** | **코오롱글로벌, 계룡건설, 동부건설, 화성산업, 태영건설, 신세계건설, 한신공영, HL D&I** |
| **Hierarchical** | **2** | **신한지주, 삼성화재, 기업은행, DB손해보험, LG유플러스, 현대해상, 코리안리** |
| **Hierarchical** | **3** | **삼화콘덴서, 삼화전기, 삼화전자** |
| **Hierarchical** | **4** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **algorithm1 - window5** | **KMeans** | **1** | **LG생활건강, CJ, 아모레G, 오뚜기, 오리온홀딩스, 대상, 동원F&B, 삼양홀딩스** |
| **KMeans** | **2** | **한올바이오파마, JW중외제약, 부광약품, 영진약품, 광동제약, 삼진제약, 일양약품, 동화약품** |
| **KMeans** | **3** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **4** | **포스코퓨처엠, 엘앤에프, 코스모신소재, 코스모화학, 송원산업, 삼익THK, 일진디스플** |
| **KMeans** | **5** | **세방전지, KG모빌리티, 넥센타이어, SNT모티브, SNT다이내믹스, SNT홀딩스, 디와이** |
| **KMeans** | **6** | **신영증권, 부국증권, 다올투자증권, 유화증권, 상상인증권** |
| **KMeans** | **7** | **CJ대한통운, 동서, 한샘, NICE평가정보, 풀무원, 한세예스24홀딩스, 벽산, 샘표, 크라운해태홀딩스, 티웨이홀딩스** |
| **KMeans** | **8** | **S-Oil, 대한유화, TKG휴켐스, 세아베스틸지주, 국도화학, 이수화학, 카프로** |
| **KMeans** | **9** | **한온시스템, 에스엘, 화신, 서연, 에스제이지세종** |
| **KMeans** | **10** | **두산에너빌리티, 한화오션, HD한국조선해양, 삼성중공업, 삼성E&A, HD현대미포, HD현대인프라코어** |
| **KMeans** | **11** | **삼성전자, SK하이닉스, 삼성SDI, LG전자, LG, 삼성전기** |
| **KMeans** | **12** | **조광피혁, BYC, 삼양통상, 대한화섬, 디씨엠, 방림, 신영와코루, 까뮤이앤씨, 대한방직** |
| **KMeans** | **13** | **신한지주, 삼성화재, 기업은행, DB손해보험, 현대해상** |
| **KMeans** | **14** | **현대건설, 대우건설, GS건설, DL, HDC** |
| **KMeans** | **15** | **NI스틸, 동양철관, DSR제강, 문배철강** |
| **KMeans** | **16** | **유한양행, 한미사이언스, 녹십자, 대웅제약, 녹십자홀딩스, 동아쏘시오홀딩스, 종근당홀딩스** |
| **KMeans** | **17** | **현대차, 기아, 현대모비스** |
| **KMeans** | **18** | **삼부토건, 코오롱글로벌, 금호건설, 계룡건설, 동부건설, 화성산업, 태영건설, 신세계건설, 한신공영, HL D&I** |
| **KMeans** | **19** | **POSCO홀딩스, 고려아연, 현대제철, 세아제강지주, 현대비앤지스틸, 동국홀딩스** |
| **KMeans** | **20** | **파미셀, 일성아이에스, 진원생명과학, 에이프로젠바이오로직스, 팜젠사이언스, 보락, 오리엔트바이오** |
| **KMeans** | **21** | **삼화콘덴서, 삼화전기, 삼화전자** |
| **KMeans** | **22** | **LG화학, LS ELECTRIC, 한화솔루션, 롯데케미칼, SKC, 금호석유, 한화, OCI홀딩스, 롯데정밀화학** |
| **KMeans** | **23** | **포스코스틸리온, 이구산업, 대창, 서원, 부국철강** |
| **KMeans** | **24** | **HMM, 쌍용C&E, 현대엘리베이, 남해화학, 현대코퍼레이션, 성신양회** |
| **DBSCAN** | **1** | **현대차, 기아, 현대모비스** |
| **DBSCAN** | **2** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **1** | **신한지주, 기업은행, DB손해보험, 현대해상** |
| **Hierarchical** | **2** | **NI스틸, 동양철관, 하이스틸, 부국철강, DSR제강, 문배철강** |
| **Hierarchical** | **3** | **삼성전자, SK하이닉스, 삼성SDI, LG전자, 삼성전기** |
| **Hierarchical** | **4** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **5** | **현대차, 기아, 현대모비스** |

**5.2. 알고리즘2 군집화 결과**

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Cluster Method** | **Cluster** | **Stocks** |
| **algorithm2 - window1** | **KMeans** | **1** | **대한항공, LS ELECTRIC, 삼성E&A, 한화솔루션, SKC, 한화, 롯데정밀화학, LX인터내셔널, SK디스커버리, 코오롱** |
| **KMeans** | **2** | **동원산업, 사조대림, 사조산업, 신라교역, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **3** | **POSCO홀딩스, LG, 한화에어로스페이스, 신세계** |
| **KMeans** | **4** | **태광산업, 고려제강, BYC, 대한화섬, 비비안** |
| **KMeans** | **5** | **삼성전자, SK하이닉스, 삼성SDI, 한국전력, 삼성전기, SK텔레콤, KT** |
| **KMeans** | **6** | **현대차, 기아, 현대모비스, 한온시스템, 한국앤컴퍼니** |
| **KMeans** | **7** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **KMeans** | **8** | **삼성화재, 기업은행, DB손해보험, 현대해상, 코리안리, 롯데손해보험, 한화손해보험, 제주은행, 흥국화재** |
| **KMeans** | **9** | **콜마홀딩스, KH 필룩스, 한국내화** |
| **KMeans** | **10** | **유한양행, 한미사이언스, 대웅, 녹십자홀딩스, JW중외제약, 동아쏘시오홀딩스, 부광약품, 종근당홀딩스, 일양약품** |
| **KMeans** | **11** | **만호제강, 성창기업지주, 방림, 대한방직, SUN&L** |
| **KMeans** | **12** | **에스엘, 화신, 서연, 에스제이지세종** |
| **KMeans** | **13** | **코웨이, 동서, 태경케미컬, 무학, 황금에스티, 인팩, 부국철강** |
| **Hierarchical** | **1** | **삼성전자, SK하이닉스, POSCO홀딩스, 삼성SDI, 한국전력, 삼성전기, SK텔레콤, KT** |
| **Hierarchical** | **2** | **코웨이, 동서, 콜마홀딩스, 태경케미컬, 에스제이지세종, KH 필룩스, 한국내화, 황금에스티, 인팩, 부국철강** |
| **Hierarchical** | **3** | **태광산업, BYC, 만호제강, 대한화섬, 성창기업지주, 방림, 대한방직, SUN&L, 비비안** |
| **Hierarchical** | **4** | **동원산업, 사조대림, 사조산업, 신라교역, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **5** | **삼성화재, 기업은행, DB손해보험, 현대해상, 코리안리, 롯데손해보험, 한화손해보험, 제주은행, 흥국화재** |
| **Hierarchical** | **6** | **현대차, 기아, 현대모비스** |
| **Hierarchical** | **7** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **algorithm2 - window2** | **KMeans** | **1** | **동원산업, 사조대림, 사조산업, 신라교역, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **2** | **SK하이닉스, 삼성SDI, LG, 삼성전기, 한화에어로스페이스, 신세계, SK네트웍스, 삼영전자, 대덕, 한국전자홀딩스** |
| **KMeans** | **3** | **현대비앤지스틸, 휴스틸, NI스틸, 동양철관, 문배철강** |
| **KMeans** | **4** | **DI동일, BYC, 만호제강, 대한화섬, 성창기업지주, 방림, 대한방직** |
| **KMeans** | **5** | **삼성전자, POSCO홀딩스, 한국전력, KT&G, SK텔레콤, KT, 한국가스공사** |
| **KMeans** | **6** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **KMeans** | **7** | **삼성화재, 기업은행, DB손해보험, 현대해상, 코리안리, 롯데손해보험, 한화손해보험, 흥국화재** |
| **KMeans** | **8** | **현대차, 기아, 현대모비스** |
| **KMeans** | **9** | **삼성E&A, GS건설, DL, HDC, HL D&I** |
| **Hierarchical** | **1** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **Hierarchical** | **2** | **삼성화재, 기업은행, DB손해보험, 현대해상, 코리안리, 롯데손해보험, 한화손해보험, 제주은행, 흥국화재** |
| **Hierarchical** | **3** | **동원산업, 사조대림, 사조산업, 신라교역, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **4** | **한국전력, KT&G, SK텔레콤, KT, 한국가스공사** |
| **Hierarchical** | **5** | **태광산업, BYC, 만호제강, 대한화섬, 성창기업지주, 방림, 대한방직** |
| **Hierarchical** | **6** | **현대차, 기아, 현대모비스** |
| **algorithm2 - window3** | **KMeans** | **1** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **2** | **두산에너빌리티, 한화오션, HD한국조선해양, 삼성중공업, 삼성E&A, HD현대미포, 두산, HD현대인프라코어, 한진중공업홀딩스** |
| **KMeans** | **3** | **한온시스템, 에스엘, 한국앤컴퍼니, 넥센타이어, 화신, 서연, 에스제이지세종, 디와이** |
| **KMeans** | **4** | **한국전력, KT&G, SK텔레콤, KT, 한국가스공사** |
| **KMeans** | **5** | **삼화콘덴서, 삼화전기, 삼화전자** |
| **KMeans** | **6** | **삼성화재, 기업은행, DB손해보험, 현대해상, 코리안리, 롯데손해보험, 한화손해보험, 흥국화재** |
| **KMeans** | **7** | **BYC, 만호제강, 대한화섬, 성창기업지주, 방림, 대한방직** |
| **KMeans** | **8** | **현대건설, 대우건설, GS건설, DL, HDC, 계룡건설, HL D&I** |
| **KMeans** | **9** | **SK가스, E1, 한국쉘석유, 태경산업, 미창석유** |
| **KMeans** | **10** | **현대차, 기아, 현대모비스** |
| **Hierarchical** | **1** | **DB손해보험, 현대해상, 롯데손해보험, 한화손해보험, 흥국화재, 진양산업, 진양폴리, 진양화학** |
| **Hierarchical** | **2** | **BYC, 만호제강, 대한화섬, 성창기업지주, 방림, 깨끗한나라, 대한방직** |
| **Hierarchical** | **3** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **4** | **한국전력, KT&G, SK텔레콤, KT, 한국가스공사** |
| **Hierarchical** | **5** | **삼화콘덴서, 삼화전기, 삼화전자** |
| **Hierarchical** | **6** | **현대차, 기아, 현대모비스** |
| **algorithm2 - window4** | **KMeans** | **1** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **2** | **신영증권, 부국증권, 한양증권, 유화증권, 상상인증권** |
| **KMeans** | **3** | **삼화콘덴서, 삼화전기, 삼화전자, 성문전자** |
| **KMeans** | **4** | **신한지주, 삼성화재, 한국전력, KT&G, SK텔레콤, 기업은행, KT, 강원랜드, 한국가스공사, 하이트진로홀딩스** |
| **KMeans** | **5** | **현대차, 기아, 현대모비스, 에스엘, 서연** |
| **KMeans** | **6** | **두산에너빌리티, 한화오션, HD한국조선해양, 삼성중공업, S-Oil, HD현대미포** |
| **KMeans** | **7** | **카카오, 대한항공, 엔씨소프트, 영원무역홀딩스, 아시아나항공, 현대지에프홀딩스, NICE평가정보, 한진, 광주신세계, 대구백화점** |
| **KMeans** | **8** | **삼성전자, SK하이닉스, 삼성SDI, LG전자, 삼성전기** |
| **KMeans** | **9** | **현대건설, 대우건설, GS건설, DL, HDC** |
| **KMeans** | **10** | **현대비앤지스틸, 대창, NI스틸, 동양철관, 서원, 한일철강, 부국철강, 문배철강** |
| **KMeans** | **11** | **코웨이, LG유플러스, 제일기획, 코리안리, 빙그레, 한섬, SBS, 한솔홀딩스** |
| **KMeans** | **12** | **POSCO홀딩스, 현대제철, 동국홀딩스** |
| **KMeans** | **13** | **유한양행, 한미사이언스, 녹십자, 대웅, 녹십자홀딩스, 동아쏘시오홀딩스, 종근당홀딩스, 환인제약, 한독** |
| **Hierarchical** | **1** | **신한지주, 삼성화재, 한국전력, KT&G, SK텔레콤, 기업은행, KT, LG유플러스, 한국가스공사, 코리안리** |
| **Hierarchical** | **2** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **3** | **삼화콘덴서, 삼화전기, 삼화전자** |
| **Hierarchical** | **4** | **삼성전자, SK하이닉스, 삼성SDI, LG전자, 삼성전기** |
| **Hierarchical** | **5** | **현대차, 기아, 현대모비스** |
| **algorithm2 - window5** | **KMeans** | **1** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **KMeans** | **2** | **한온시스템, 에스엘, 화신, 서연, 에스제이지세종** |
| **KMeans** | **3** | **휴스틸, NI스틸, 동양철관, 대호에이엘, 하이스틸, 한일철강, 부국철강, DSR제강, 문배철강** |
| **KMeans** | **4** | **신한지주, 삼성화재, 한국전력, SK텔레콤, 기업은행, KT, LG유플러스, 강원랜드, 한국가스공사** |
| **KMeans** | **5** | **삼성전자, SK하이닉스, 삼성SDI, LG전자, LG, 삼성전기** |
| **KMeans** | **6** | **현대차, 기아, 현대모비스, 한국앤컴퍼니** |
| **KMeans** | **7** | **삼화콘덴서, 삼화전기, 삼화전자** |
| **Hierarchical** | **1** | **신한지주, 삼성화재, 한국전력, SK텔레콤, 기업은행, KT, LG유플러스, 강원랜드, 한국가스공사, 코리안리** |
| **Hierarchical** | **2** | **휴스틸, NI스틸, 동양철관, 대호에이엘, 하이스틸, 한일철강, 부국철강, DSR제강, 문배철강** |
| **Hierarchical** | **3** | **삼화콘덴서, 삼화전기, 삼화전자** |
| **Hierarchical** | **4** | **동원산업, 사조대림, 사조산업, 신라교역, CJ씨푸드, 사조오양, 한성기업, 동원수산** |
| **Hierarchical** | **5** | **현대차, 기아, 현대모비스** |
| **Hierarchical** | **6** | **삼성전자, SK하이닉스, 삼성SDI, LG전자, 삼성전기** |

**5.3 전략 벡테스트**







X. Reference

[1] Simão Moraes Sarmento, and Nuno Horta. “Proposed Pairs Selection Framework” A Machine Learning based on Pairs Trading Investment Strategy (2020): 21-35 [[LINK](https://link.springer.com/chapter/10.1007/978-3-030-47251-1_3)]

[2] Hubel, David H., and Torsten N. Wiesel. ＂Receptive fields, binocular interaction and functional Architecture in the cat＇s visual cortex.＂ The Journal of physiology 160.1 (1962): 106. [[LINK](https://link.springer.com/chapter/10.1007/978-3-030-47251-1_3)]

[8] Harlacher M (2016) Cointegration based algorithmic pairs trading. PhD thesis, University of st.Gallen

[9] Moraes Sarmento, S., Horta, N. (2021). Proposed Pairs Selection Framework. In: A Machine Learning based Pairs Trading Investment Strategy. SpringerBriefs in Applied Sciences and Technology(). Springer, Cham. https://doi.org/10.1007/978-3-030-47251-1\_3

본 자료는 경북대학교 금융 데이터분석학회 DART의 제작물로서 모든 저작권은 작성한 학회의 조사분석담당자 본인에게 있습니다. 본 자료는 학회의 동의 없이 어떠한 경우도 변형, 복제, 배포, 전송, 대여할 수 없습니다. 본 자료에 수록된 내용은 학회 및 조사분석담당자가 신뢰할 만한 분석 및 자료로부터 얻은 것이나, 본 학회는 그 정확성과 완전성을 보장할 수 없습니다.