

2단계 k-평균 군집화를 활용한 한류콘텐츠 기업 주가 예측 연구

김정우
강릉원주대학교 경제학과 교수

A Study On Predicting Stock Prices Of Hallyu Content Companies Using Two-Stage k-Means Clustering

Jeong-Woo Kim
Professor, Dept. of Economics, Gangneung Wonju National University

요약 본 연구는 기존의 k-평균 군집화를 활용한 2단계 k-평균 군집화 방법을 사용하여 한류콘텐츠 기업들의 주가가격을 예측함으로써 본 기법이 예측성능을 개선할 수 있음을 보이고자 하였다. 이를 위하여 본 연구는 2단계 k-평균 군집화의 알고리즘을 소개하고, 다양한 머신러닝 기법들과의 예측값 비교를 통하여 본 기법의 예측성능을 검증하였다. 본 기법은 기존의 k-평균 군집화로부터 얻어진 군집들 중에서 예측 대상에 근접한 군집을 추출하고 이 군집에 k-평균 군집화 방법을 다시 적용하여 실제 값에 보다 근접한 군집을 탐색하는 방식이다. 본 기법을 한류콘텐츠 기업들의 주가 시계열 자료에 적용한 결과, 다른 머신러닝 기법의 예측값들보다 실제 주가가격에 근접한 예측값을 나타내어, 기존의 k-평균 군집화 방법보다 개선된 예측성능을 보였다. 또한, 본 기법은 상대적으로 적은 크기의 군집을 사용함에도 불구하고 비교적 안정적인 예측값을 나타내었다. 이에 따라, 2단계 k-평균 군집화 기법은 예측의 정확성과 안정성을 동시에 개선할 수 있으며, 소규모 자료에도 유용할 수 있는 새로운 군집화 방식을 제시했다고 볼 수 있다. 향후에는 본 기법을 발전시켜 대규모 자료에도 적용하는 방안을 검토하는 연구가 요구된다.

주제어 : 군집화, 머신러닝, 과적합, 예측, 시계열 자료

Abstract This study shows that the two-stage k-means clustering method can improve prediction performance by predicting the stock price. To this end, this study introduces the two-stage k-means clustering algorithm and tests the prediction performance through comparison with various machine learning techniques. It selects the cluster close to the prediction target obtained from the k-means clustering, and reapplies the k-means clustering method to the cluster to search for a cluster closer to the actual value. As a result, the predicted value of this method is shown to be closer to the actual stock price than the predicted values of other machine learning techniques. Furthermore, it shows a relatively stable predicted value despite the use of a relatively small cluster. Accordingly, this method can simultaneously improve the accuracy and stability of prediction, and it can be considered as the new clustering method useful for small data. In the future, developing the two-stage k-means clustering is required for the large-scale data application.

Key Words : Clustering, Machine learning, Overfitting, Prediction, Time-series data

*This study was supported by 2021 Academic Research Support Program in Gangneung-Wonju National University.

*Corresponding Author : Jeong-Woo Kim(kurtkim@gwnu.ac.kr)

Received April 19, 2021

Accepted July 20, 2021

Revised June 25, 2021

Published July 28, 2021

1. 서론

전통적으로 예측 연구는 기존 이론에 근거하여 변수들의 인과관계 검증 후 이를 기반으로 예측을 시도하였다. 가장 바람직한 경우는 검증된 인과관계에 기초하여 예측 연구를 수행할 시 예측정확성이 높게 나타나는 것이다. 하지만 특정 모델이 주어진 데이터의 인과관계 설명력만을 높이고자 하면, 새로운 데이터에 모델의 적용가능성은 낮아지고 이에 따라 예측정확성도 낮아지는 과적합 문제가 발생할 수 있다[1].

실용적인 측면에서의 예측 연구는 산업계에 적용되는 것이 경우가 많았으며, 특히 기업들의 주가 예측에 대해서는 수많은 정량적 연구가 이루어져 왔고 이러한 연구들은 기업의 특성에 따라 구분하여 볼 수 있다. 예를 들어, 규모가 크고 업력이 오래된 제조업 업체의 경우에는 기존의 전통적인 계량모형을 재무적 정보에 적용하여 주가를 예측할 수 있으며 재무제표, 공시자료 등과 같은 신뢰성 높은 정보를 기반으로 예측하는 경우가 많았다[2-4]. 이러한 예측 방식은 안정성은 높으나 사용되는 정보의 최신성이 낮으므로 비교적 최신의 정보를 사용하는 접근 방식이 요구된다고 볼 수 있다. 한편, 코스닥 기업에 속하는 업력이 비교적 짧고 규모가 상대적으로 작은 회사들의 경우에는 예측에 필요한 상기와 같은 정보가 충분치 않거나 정보 접근성이 적은 경우가 대부분으로 주가 예측이 어렵고 변동성 또한 크다고 볼 수 있다. 이에 따라, 개인투자자들은 호재, 소문 등 정보의 최신성은 높으나 신뢰성은 비교적 낮은 정보에 기반하여 투자를 하여 피해를 입는 경우가 발생하기도 하였다. 과거, 벤처 붐 이후 투자자들의 피해[5]와 최근 바이오 기업들의 주가 급변동에 따른 투자자들의 손실[6]은 이러한 점을 반영하는 사례들이다. 특히, 최근 신용대출로 주식에 투자하는 젊은 층들이 많아지고 있다는 상황[7]은 이러한 개인투자자의 피해가 앞으로 더 커질수도 있다는 점을 짐작하게 한다고 볼 수 있다. 이에 따라, 코스닥 기업들과 같이 비교적 소규모 회사들의 주가를 합리적이고 안정적으로 예측하는 접근법이 필요하다고 볼 수 있으며, 정보의 최신성이 높은 비재무적 정보를 머신러닝 기법으로 분석하는 것은 또 다른 접근 방식이 될 수 있다[8-10]. 이러한 점들을 고려할 시, 재무적 정보에 기반한 기존의 계량 모델 외에 정보의 최신성이 높은 비재무적 정보를 활용한 머신러닝 분석은 개인투자자들의 안정적인 주가 예측을 도모하여 투자결정에 합리성을 제고하는데 필요한 과제라고 볼 수 있다.

상기 요인들을 고려하여 본 연구는 기존의 재무적 정보 기반 계량모델 중심의 주가 예측 연구들과 달리 정보의 최신성이 높은 비재무적 정보를 사용하는 머신러닝 기법을 사용하는 방법론적 차별성을 우선 확보하고자 하였다. 또한, 기존의 비재무적 정보를 이용한 연구들이 다양한 업종의 비교적 규모가 큰 코스피 상장 기업들을 분석한 것과 달리[9,10], 본 연구는 코스닥 기업 중 비재무적 정보의 영향력이 상대적으로 크다고 볼 수 있는 한류 콘텐츠 기업을 대상으로 삼아 연구 대상에서의 차별성을 지니고자 하였다.

한류콘텐츠 산업은 산업특성과 우리나라의 자원구조를 고려할 시 중요성이 높은 산업분야라고 볼 수 있다. 특히, 한류의 중심에 있는 k-pop은 한국의 한류콘텐츠 산업의 수출시장을 선도하고 있다고 해도 과언이 아니다. k-pop은 음악의 완성도뿐만 아니라 안무, 의상, 마케팅 등 음악 외적인 것들이 중요한 요소로 작용하므로 다양한 요소들이 성패에 영향을 주는 시장이라고 할 수 있다. k-pop 시장에는 대기업 자본도 많이 유입되어 있지만 소속 음악가들을 고려할 시 대형 엔터테인먼트사 S사, Y사, J사 등 3개 회사가 시장에서 가지는 지위는 높다고 볼 수 있다. 이 3개 회사들은 코스닥에 상장되어 있으며, 관련된 온라인 상의 뉴스, SNS 등의 양을 감안하면 비재무적 요소가 주가에 영향을 줄 수 있다고 볼 수 있다[8]. 또한, 상기 서술한 비재무적 정보가 주가에 주는 영향력을 고려하면, 한류콘텐츠 회사 주가는 제조업 등 전통적인 산업이 회사들보다 변동성이 클 수 있으며, 이에 따라 주식 가격의 예측이 쉽지 않다는 것을 가늠할 수 있다.

일반적으로 주식 가격의 예측에는 주가와 관련된 변수를 포함한 회귀분석 또는 주가 데이터만을 고려한 시계열 분석 등이 많이 사용되었다. 현재 값과 과거 값의 관계를 선형으로 분석한다는 점에서 시계열 분석도 회귀분석의 기본적 방법론을 따르고 있다고 볼 수 있다. 회귀분석은 특정 이론에 근거한 변수들 사이의 일정한 관계를 설정하고 이 관계에서 벗어나는 정도를 오차항으로 간주하여 연구대상의 변수 간 관계를 분석한다. 반면에, 변수들 사이의 관계에 대한 사전 지식없이 데이터의 패턴만을 분석하여 변수의 미래 값을 예측하는 접근방법도 있는데, 대표적으로 머신러닝 기법은 회귀분석에서의 오차항으로 간주되는 부분들도 데이터의 패턴을 분석하는데 사용하는 방법이다.

물론, 머신러닝 기법들이 회귀분석과 전혀 다른 접근 방식을 취하는 것은 아니다. 오히려, 많은 머신러닝 기법

들이 선형회귀분석 모델을 기초로 개발된 것이 사실이다. LASSO(Least absolute shrinkage and selection operator)는 직접적으로 선형회귀분석 모델을 토대로 만들어졌으며, 주성분분석(Principal Component Analysis, PCA) 등의 방법도 변수들의 선형관계로부터 파생된 머신러닝 기법들이다. 다만, 이러한 방법들은 변수들의 관계를 사전에 설정하지 않고 변수들의 계수를 조정하기 쉬운 함수식으로 설정함으로써, 데이터 기반에서 변수들 사이의 관계를 파악하는 것이다.

한편, 머신러닝 기법 중 변수들 사이의 관계를 파악하는 데에 중점을 두는 상기 방법들과 달리 주어진 데이터로부터 새로운 하위 샘플을 추출하여, 이 하위 샘플만을 이용하는 머신러닝 기법들이 있다. 이러한 방법들 중에는 k-최근접 이웃법(k-Nearest Neighbor), k-평균 군집화(k-means clustering, kmeans) 등이 대표적이다. 이러한 머신러닝 기법들은 변수들 사이의 관계를 파악하는데 중점을 두기보다는, 예측 대상과 근접한 하위 자료들의 유사성에 근거하여 예측 작업을 수행하는 것이다. kNN은 예측 대상에 근접한 자료들만을 선택한 후 이 선택 자료들의 평균 등을 예측치로 간주하는 기법이다. kNN이 예측대상의 근방 자료만을 기반으로 예측하는 방법이라면, k-평균 군집화는 주어진 전체 데이터를 적절한 개수의 하위 군집들로 구분하고 예측 대상과 유사도가 가장 높은 하위 군집을 기반으로 예측결과를 산출하는 방법이다. 그러므로, kNN과 달리 k-평균 군집화는 주어진 자료 전체의 특징을 고려하는 장점이 있으며, 자료의 개수가 많은 경우에도 쉽게 적용이 가능하다.

k-평균 군집화는 관측치 간의 유사도를 기준으로 하위 군집을 결정하므로, 유사도를 어떻게 정의하느냐에 따라서 다양한 방식의 k-평균 군집화 기법들이 존재한다 [11]. 또한, 다른 머신러닝 기법들과 마찬가지로 k-평균 군집화도 분류 목적에 많이 활용되어 왔기 때문에 상기 k-평균 군집화 기법들도 유사도에 기반하여 분류 정확도를 높이기 위해 제시된 경우가 많았다. 반면에 분류 작업의 레이블(Label)이 범주형 변수인 것과 달리, 예측 작업의 변수는, 특히 시계열 자료(Time-series data)의 경우는, 연속형 변수이므로 보다 세밀하게 구분된 하위 군집이 생성 가능하다고 볼 수 있다. 하지만, k-평균 군집화의 초기 k값은 보통 10개 내외로 설정되어 최적의 k값을 탐색하는 경우가 대부분으로 [12], 보다 세분화된 하위군집을 탐색하는 데에는 제한적이었다. 이에 따라, 본 연구는 전체 데이터로부터 보다 세밀하게 구분된 하위 군집을 생성해내기 위하여 기존의 k-평균 군집화 기법을 변

형한 방법을 사용하여 예측 연구를 수행하고자 하였다. 이러한 예측 기법은 코스닥 시장의 한류콘텐츠 회사와 같이 비재무적 요소 등에 영향에 따라 예측 시점에서의 주가의 변동성이 큰 경우에 적합한 접근 방식이라고 볼 수 있다. 이에 따라, 본 연구는 기존의 k-평균 군집화 기법을 변형한 방법을 비재무적 정보 분석에 적용하여 한류콘텐츠 회사의 주가를 예측하는 것을 연구 목적으로 한다.

2. 이론적 배경

일반적으로 주식 가격 예측과 관련된 연구는 오랜 기간 동안 많은 분야에서 이루어졌다. ARIMA 모형 등을 사용하여 코스피200 지수를 예측한 연구 [13], AR, GARCH 등의 시계열 모형으로 주가지수 수익률을 예측한 연구 [14] 등은 전통적 계량모형을 사용한 사례들이다. 또한, 기업의 영업이익과 주가변동의 관계를 일반회귀모형을 통하여 분석한 연구 [2], 주당순이익, 주당순자산 등의 정보를 활용하여 주가를 예측한 연구 [3], 회계정보의 정정이 주가변동에 미치는 영향을 분석한 연구 [4] 등은 재무정보를 기반으로 한 주가 예측 연구 사례들이다. 하지만, 이러한 연구들은 신뢰성이 높은 재무 정보의 습득이 용이하고 재무정보 기반의 계량모형이 비교적 장기간에 걸쳐 유효한 코스피 기업들을 대상으로 이루어진 경우가 대부분이다. 이에 따라, 기업 여건이 상이하여 비재무적 정보의 영향력이 상대적으로 큰 소규모 기업의 주가를 예측하는 경우에는 다른 분석적 접근이 보완될 필요가 있다고 할 수 있다.

한편, 비교적 소규모인 코스닥 기업들의 주가 등락 예측을 딥러닝 기법으로 분석한 연구 [15], 코스닥 기업들의 회계정보를 기업가치모형으로 분석하여 주가를 예측한 연구 [16] 등의 주가 예측 연구들이 기존에 이루진 바 있으나, 업종을 구분하여 이루어지지는 않았으며, 특히 본 연구의 대상인 한류콘텐츠 기업들에 대한 주식 가격 예측과 관련된 연구는 많지 않은 편이다. 이것은 한류콘텐츠 산업이 비교적 후발 산업인 점도 작용하지만 재무 정보와 같은 요소 외에도 뉴스, 이슈 등의 비재무적 요소 등이 주식 가격에 영향을 크게 끼치므로 연구대상으로 주목받지 못한 부분도 있다고 하겠다. 한편, 사회의 이슈화 정도와 관련된 뉴스, SNS, 키워드 검색 등과 주식가격 간의 관계를 분석한 연구는 기존에 이루어진 바 있다. 기업의 페이스북에 대한 좋아요 수와 주식 가격의 상관

성을 밝힌 연구[17], SNS 및 뉴스 등의 온라인 정보를 이용하여 주가를 예측한 연구[9,18] 등은 온라인 상의 비재무적 정보가 주식 가격에 영향을 준다는 것을 잘 보여 주는 사례라고 하겠다. 하지만, 이러한 연구들은 기존에 존재하는 계량모델 또는 머신러닝 기법들을 주어진 자료에 적용하는 수준이었다.

상기와 같이 기존 연구를 검토한 결과, 재무정보 접근성과 활용성이 상대적으로 낮은 소규모 한류콘텐츠 기업의 주가 예측에 재무정보 기반의 계량모델을 일률적으로 적용하기에는 다소 무리가 있으며, 정보의 최신성이 높은 비재무적 정보를 활용한 주가 예측 방법이 보완책이 될 수 있다고 보인다. 또한, 기존의 계량모델, 머신러닝 기법을 단순히 적용하기 보다는 새로운 예측 기법을 개발하여 주가 예측을 하는 것은 독창적인 기법으로 예측 정확성을 개선한다는 점에서 기존 연구들과 본 연구가 차별화되는 중요한 요소라고 하겠다.

본 연구에서 기초적인 방법론으로 사용되는 k-평균 군집화 방법이 데이터 마이닝 분야에 도입된 이래로[19], 이 방법은 다양한 분류 및 예측 연구에서 사용되었다. 학생들의 학업성취도 예측[20], 유전자 정보로부터 암의 종류 예측[21], 여행에 소요되는 여행시간 예측[22], 심장마비에 영향을 주는 다양한 요소를 고려하여 심장마비를 예측[23], 기업의 성과예측[24] 등 k-평균 군집화 방법은 자연과학 및 사회과학 분야에서 광범위하게 사용되었다. 이는 데이터 기반의 분류 및 예측 작업이 관측치 간의 유사도를 기반으로 이루어지는 경우가 많으므로, 상기 연구들과 같이 k-평균 군집화 방법의 적용이 유연한 것이라고 볼 수 있다. 또한, k-평균 군집화 방법은 비교적 단순한 수학적 모델을 따르고 있으므로, 다양한 방식으로 변경되거나 다른 머신러닝 기법과 함께 사용되기도 하였다. k-평균 군집화와 결정 나무(Decision tree) 기법을 결합한 방법으로 당뇨병 발병을 예측[25], 주성분분석과 k-평균 군집화를 사용하여 유방암을 예측[26], k-평균 군집화와 신경망 모형(Neural network)을 결합한 기법을 통하여 시간당 태양복사열을 예측[27]한 일련의 연구들은 이러한 연구 사례들이라고 볼 수 있다.

k-평균 군집화 방법은 정확한 분류와 예측을 위하여 주어진 데이터 전체를 해석가능한 특정한 함수형태를 가정하기보다는, 유사한 하위 군집을 우선 추출하는 방식을 취하고 있다. 이에 따라, 부분 군집들의 유사도에 따른 분류 및 예측을 수행하는 k-평균 군집화 방법은 상기 연구들과 같이 의학, 생물학 분야 등에서 활발히 사용되었다. 또한, 특정한 함수형태를 가정하지 않는다는 점에

서 k-평균 군집화 방법은 다양한 머신러닝 기법들과 결합되어 실증연구에 적용될 수 있었다.

아울러, 본 연구에서 제안하는 2단계 k-평균 군집화 방법과 관련성이 높은 군집화 방법들이 과거에 제안된 바 있다. 이른바 이중 k-평균 군집화 방법(Double k-means clustering)은 k-평균 군집화 방법을 각 군집의 평균과 분산이 수렴할 때까지 반복 적용하여 예측오류를 감소된다는 것을 보여준 연구이며[28], 분석 대상에서 무작위로 k개의 군집을 선택하고 군집의 중심이 안정화되는 시점까지 군집화를 수행하는 알고리즘인 반복 k-평균 군집화 방법(repeated k-means clustering)은 뇌의 영역 간 연결성을 분석하는데 적용되어 분류 결과의 안정성을 개선하는 것으로 나타났다[29]. 또한, k-평균 군집화의 분석 결과가 초기 k값에 큰 영향을 받는다는 점에 착안하여 군집화 작업 후의 각 군집들의 개수와 크기를 조정하여 적절한 k값을 다시 탐색하는 중첩 k-평균 군집화 방법(repeated k-means clustering)[30], 주어진 자료의 일부분만을 첫 단계에서 사용하여 군집의 중심을 탐색하고 다음 단계에서 첫 단계에서 추출된 중심을 기반으로 정확한 중심의 위치를 탐색하는 방식으로 군집화 연산 시간을 단축시킬 수 있는 2단계 군집화(Two stage clustering) 기법[31], EM(Expectation-maximization) 알고리즘으로 초기 k값을 도출한 후 군집화를 수행하는 2층계 군집화(Two step clustering) 기법[32] 등이 관련 연구로써 수행된 바 있다.

상기 방법들은 대체로 기존의 k-평균 군집화 방법이 연구자가 임의로 설정한 초기 k값에 의존하는 문제점에 착안하여[12,33], 주어진 자료 전체를 대상으로 k-평균 군집화 방법을 반복 적용하는 방식으로 군집의 안정성[28,29]과 연산 효율성[31]을 도모한 방법들이라고 볼 수 있다. 하지만, 상기 방법들은 주어진 전체 자료를 군집화 대상으로 하고 있어 불규칙한 시계열 자료를 예측하는 데에는 다른 방법이 필요하다고 볼 수 있다. 즉, 불규칙한 시계열 자료가 주어진 경우에는 자료 전체보다는 예측 시점과 근접한 구간의 특성을 더 많이 고려할 필요가 있는 경우에는 상기 방법들과는 다른 접근 방식이 유효할 수가 있는 것이다. 본 연구에서는 이러한 방법론적인 측면에 착안하여, 2단계 방식의 k-평균 군집화 방법을 사용하여 한류콘텐츠 기업의 주식 가격을 예측해보고자 한다. k-평균 군집화 방법과 관련된 상기 연구들을 정리하면 Table 1과 같다.

Table 1. Related works

Type	Author	Subject
Application	Park et al.(2009)	KOSPI200 Index Prediction using ARIMA, etc
	Kim & Klm(2009)	KOSPI Yield prediction using AR, GARCH, etc
	Kwon et al.(2010)	Firm's operating profit and stock price volatility
	Heo & Yang(2015)	Stock price prediction using EPS, BPS, etc
	Park et al.(2015)	Effect of the corrective disclosure on stock price
	Song & Lee(2018)	KOSDAQ firm's stock price using deep learning
	Song(2002)	Stock price prediction based on the accounting information of KOSDAQ firms
	Oyelade et al.(2010)	Predicting academic achievement by k-means clustering
	Nidhees et al.(2017)	Prediction of cancer type using k-means clustering
	Nath et al.(2010)	Application of k-means clustering to predict travel time
Methodology	Shinde et al.(2015)	Predicting heart attack using the k-means clustering
	Lee(2011)	Application of k-means clustering to develop a firm's performance prediction model
	Chen et al.(2017)	Combining k-means clustering and decision tree
	Jamal et al.(2018)	Combine principal components analysis and k-means clustering
	Benmouzaa & Cheknaneb(2013)	Combining k-means clustering and neural network models
	Huh(2000)	Double k-means clustering to enhance cluster stability
	Nanetti et al.(2009)	Iterative k-means clustering to obtain converging cluster centers
	Hassan(2018)	Repeated k-means clustering using adjusted clusters
	Salman et al.(2011)	Two-step clustering for computational efficiency
	Chayangkoon & Srivihok(2016)	Two-step clustering using EM algorithm

3. 연구방법

k-평균 군집화 방법의 기본적인 수식은 식 (1)과 같다.

$$S = \arg \min_S \sum_{i=1}^K \sum_{x \in S_i} \|x - \mu_i\|^2 \quad (1)$$

여기서 x 는 주어진 데이터, $S = \{S_i, i = 1, \dots, K\}$ 는 군집들의 집합, μ_i 는 각 군집의 평균값이며, K 의 값은

데이터의 특성을 고려하여 정할 수 있다. 각 군집들은 각 군집에 속한 관측치가 각 군집의 평균까지의 거리가 최소화되도록 선택된다. 즉, k-평균 군집화 방법은 군집 내 (Within-cluster) 분산을 최소화하는 방식을 따르고 있으며, 이에 따라 관측치와 군집의 평균 간의 거리는 유클리디안(Euclidean) 거리이다. 물론, 데이터의 속성에 따라 다양한 거리척도가 사용되어 분류 또는 예측에서 다른 결과가 도출될 수 있다[33,34]. k값은 연구마다 임의로 주어지거나 10개 내외에서 탐색되는 경우가 많다. 하지만, 데이터의 크기가 크거나 하위 군집 내의 변동이 큰 경우에는 이와 같은 k값의 초기 설정은 정확한 예측에 적절하지 않을 수 있다. 반면에, k의 초기값을 큰 값으로 설정하면, 연산 규모가 커지므로 비효율적이다[35]. 또한, k값이 너무 큰 경우에는 예측의 목표값과 근접한 군집뿐만 아니라, 목표값과 거리가 먼 군집들도 식 (1)의 최적화 문제에 포함해야 하므로 목표값의 정확한 예측 정확성이 낮아질 수 있다. 이에 따라, 본 연구에서는 군집을 세분화하면서 예측 목표값과 근접한 군집만을 고려하는 2단계 k-평균 군집화 방법(Two stage k-means clustering, kmeanscl)을 사용하여 예측 정확성을 확보하고자 하였다. 이 방법을 활용한 예측 방법은 다음과 같다.

우선 주어진 데이터 개수를 N 이라고 두고, $N+1$ 번째의 값을 예측하는 문제를 고려해보자. (1)을 통하여 N 크기의 데이터는 총 K 개의 군집으로 나누어질 수 있다. 여기서 $N+1$ 번째와 가장 가까운 군집, 즉 시계열 자료 예측의 문제에서는 K 번째의 군집을 새로운 군집화 대상 S_K 로 간주하자. 그러면, S_K 의 군집화는 식 (2)와 같다.

$$C = \arg \min_C \sum_{i=1}^l \sum_{z \in C_i} \|z - m_i\|^2 \quad (2)$$

여기서 z 는 S_K 내에서의 데이터, $C = \{C_i, i = 1, \dots, l\}$ 는 새로운 군집들의 집합, m_i 는 각 군집의 평균값이다. 새로운 군집들 C_i 중에서 예측 목표인 $N+1$ 번째 자료와 가장 근접한 군집을 C_o 라고 하면, 예측값은 이 C_o 를 활용하여 얻을 수 있다. 이와 같이 얻어진 군집 C_o 는 전체 데이터를 처음부터 최대 Kl 개의 군집으로 나누어 얻어진 군집과 유사하다. 하지만, 2단계의 군집화를 사용하면 총 $(K+l)$ 개의 군집만을 고려하여 C_o 를 얻을 수

있다. 또한, C_0 은 1단계 군집화에서 얻은 군집들 중에서 예측 목표와 가장 가까운 S_K 군집내에서 얻어지므로 예측 정확성을 높이는데 유용할 수 있다. 이러한 방식으로 얻어진 군집은 장기적인 데이터의 패턴을 고려하기 보다는 예측 시점에 근접한 데이터들만을 고려하므로 뉴스, 이슈 등의 비재무적 요소에 의해 불규칙적으로 변동하는 콘텐츠 회사의 주가 예측에 적절할 것으로 볼 수 있다. 이상의 2단계 k-평균 군집화 방법을 정리하면 Fig. 1과 같다.

Algorithm: Two stage k-means clustering	
1	Set the initial number of cluster k for a given dataset $X = \{x x \in R\}$
2	Select the class $S = \{S_i, \text{mean}(S_i) = \mu_i, i = 1, \dots, k\}$ as follows
3	$S = \text{argS} \min \sum_{i=1}^k \sum_{x \in S_i} \ x - \mu_i\ ^2$
4	Choose the cluster S_K nearest to the prediction target
5	Do step 3 for S_K to obtain the subclass, $C = \{C_i, i = 1, \dots, l\}$
6	Determine the subcluster C_0 nearest to the prediction target
7	Return the prediction value based on C_0

Fig. 1. Two stage k-means clustering algorithm

본 연구는 2단계 k-평균 군집화 방법의 예측 성능을 확인하기 위하여 기본적인 최소자승법(OLS) 및 다양한 머신러닝 기법과 비교를 하였다. 머신러닝 기법 중 LASSO는 일반적인 선형회귀 분석 시 발생할 수 있는 과적합(Overfitting) 문제를 완화할 수 있는 방법으로 알려져 있는데[1]. 이는 LASSO를 통한 회귀계수는 조정 파라미터(Tuning parameter)를 통하여 일반적인 선형회귀에 의한 회귀계수보다 적게 추정되기 때문이다. 이를 통하여, LASSO는 주어진 데이터에 대한 설명력만을 고려함으로써 발생할 수 있는 과적합 문제를 완화할 수 있다. OLS나 LASSO가 특정 회귀 모형을 추정하는데 중점을 둔다면, 서포트벡터머신(Support Vector Machine, SVM)은 독립변수들을 활용하여 주어진 종속변수를 최대한 잘 분류하는데 중점을 두는 머신러닝 기법이다. SVM은 독립변수들을 구분하는 함수식을 설정하고, 이에 따라 종속변수를 얼마나 잘 분류하고 예측하는지에 따라 함수식을 결정하는 방법이다. SVM에서 쓰이는 함수식은 주로 독립변수 간의 거리함수로 복잡한 모델을 설정할

필요가 없다. 본 연구에서는 SVM을 회귀모형에 적용한 서포트벡터회귀(Support vector regression)를 사용하였다[36]. 신경망(Neural Network, NN) 모형은 데이터의 특성을 고려하거나 특정한 함수식을 설정하지 않고 독립변수를 반복적으로 조합하여 새로운 변수를 생성해내는 방법으로 예측성능을 높이는 기법이다. 즉, 신경망 모형은 독립변수에 가중치를 부여하여 은닉층(Hidden Layer)을 생성하고, 이 은닉층들을 종속변수에 적합시켜 예측성능이 가장 높아지는 시점까지 이 과정을 반복하는 것이다. 또한 신경망 모형은 독립변수를 종속변수에 적합한 과정에서 발생가능한 과적합 문제를 역전파(Backpropagation)를 통하여 완화한다. 역전파는 신경망 모형을 통한 예측값과 실제값의 차이를 상기 신경망 모형 과정에 반영하여 가중치 등을 수정하는 방법으로 신경망 모형이 높은 예측력을 지니는데 중요한 역할을 한다[22]. 마지막으로 2단계 k-평균 군집화 방법이 기존의 k-평균 군집화 방법보다 높은 예측성능을 보이는지 확인하기 위하여 k-평균 군집화 방법을 비교 예측기법으로 포함하였다.

4. 분석결과

본 연구에서는 S사, Y사, J사의 코스닥 주식 가격 시계열 자료에 2단계 k-평균 군집화 방법 및 다양한 예측 기법을 적용하여 주식 가격을 예측하였다. 주가 예측에 사용된 비재무적 정보로서는 기존의 관련 연구들이 뉴스, 온라인 검색 수 등을 주가 연구에 적용한 것을 고려하여[8,37], 각 콘텐츠 회사의 이름과 대표 음악그룹명의 검색 수 그리고 한류와 k-pop이라는 키워드 검색 수를 사용하였다(국내 포털사이트 N사 데이터를 사용). 또한, 콘텐츠 산업에 대한 수요를 반영하기 위하여 소비자동향지수(한국은행) 중 교양·오락·문화생활비 지출전망지수, 주가의 금리민감도를 고려한 시장금리(KORIBOR 3개월), 전반적 시장상황을 반영하는 코스닥지수도 포함하였다. 자료 수집기간은 2018년 1월 2일부터 2020년 12월 30일까지이고, 예측기간은 2019년 10월 15일부터 2020년 12월 30일까지이며, 자료의 계절 조정을 위하여 원 시계열 자료에서 계절성분을 제외한 후 분석에 사용하였다.

데이터에 대한 기술통계량은 Table 2와 같다. 우선 종속변수인 주식 가격(Price)을 보면 평균값보다 표준편차보다 커서 주식 가격의 변동성이 크게 나타나 한류콘

텐츠 기업의 불규칙한 주식 가격 특성을 반영한다고 볼 수 있다. 독립변수 중에서 일반적인 성격의 키워드인 k-pop과 한류의 경우는 평균이 표준편차보다 큰 반면, 일반성이 보다 적은 각 회사명과 대표 음악그룹명은 표준편차가 평균보다 대부분 큰 경향을 나타내어 비교적 큰 변동성을 보여주고 있다.

Table 2. Descriptive statistics

(Unit: Won, %)

Firm	Variable	MEAN	STD	MIN	MEDIAN	MAX
S	Price	37329	7635	16700	37400	56100
	ss	7.74	7.41	0.67	5.86	100.00
	rvs	1.65	5.00	0.36	0.83	100.00
Y	Price	34736	8380	19450	33300	58700
	ys	2.75	6.36	0.00	1.52	100.00
	bls	2.48	4.64	0.76	1.63	100.00
J	Price	26155	6581.85	12800	24450	43250
	js	10.08	9.09	0.00	8.61	100.00
	twvs	2.21	4.74	0.63	1.51	100.00
Comm on	kpops	6.46	4.69	2.60	6.01	100.00
	halus	44.40	12.35	15.48	44.08	100.00
	Csi	87.32	5.91	74	90	94
	Kori	1.41	0.43	0.57	1.64	1.9
	Kos	747.25	98.46	428.35	742.32	968.42

* STD(Standard deviation), ss(Firm S search), ys(Firm Y search), js(Firm J search), kpops(k-pop search), halus(hallyu search), rvs(Firm S' artist search), bls(Firm Y's artist search), twvs(Firm J's artist search), Csi(Consumer Sentiment Index), Kori(KORIBOR), Kos(KOSDAQ), the number of search is scaled by percentage.

본 연구에서는 2단계 k-평균 군집화 방법과 다른 예측기법들의 예측성능을 비교하기 위하여 평균절대비오차 (Mean Absolute Percentage Error, MAPE)를 사용하였다. MAPE는 식 (3)과 같다.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - f_i|}{y_i} \times 100 \quad (3)$$

여기서, N 은 예측 수, y_i 는 실제값, f_i 는 각 기법들의 예측값임.

Table 3은 각 예측기법들의 MAPE값을 회사별로 나타낸 것이다. 2단계 k-평균 군집화 방법(kmeanscl)은 모든 회사에서 가장 낮은 MAPE값을 보여주어 예측성능이 높은 것으로 나타났다. k-평균 군집화 방법도 비교적 낮은 MAPE값을 나타내고 있는데 이는 주식 가격의 변

동성이 커서 선형모형을 사용하고 있는 OLS나 LASSO 기법의 예측 정확성이 낮다는 점과 연관이 있다고 볼 수 있다.

Table 3. MAPE comparison

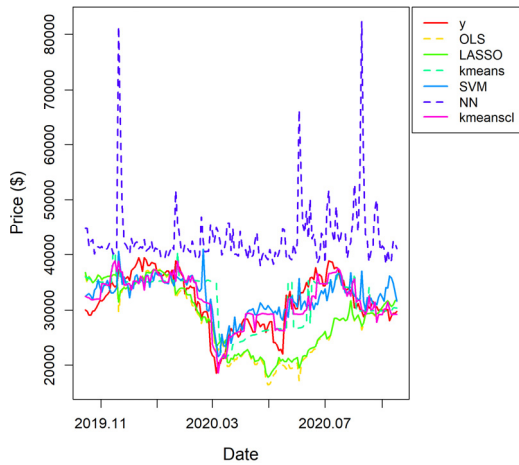
(Unit: %)

Firm	OLS	LASSO	kmeans	SVM	NN	kmeanscl
S	14.96	14.16	8.31	6.72	38.7	3.89
Y	24.32	24	9.14	8.63	22.69	4.58
J	25.1	24.76	6.62	8.37	17.52	3.46

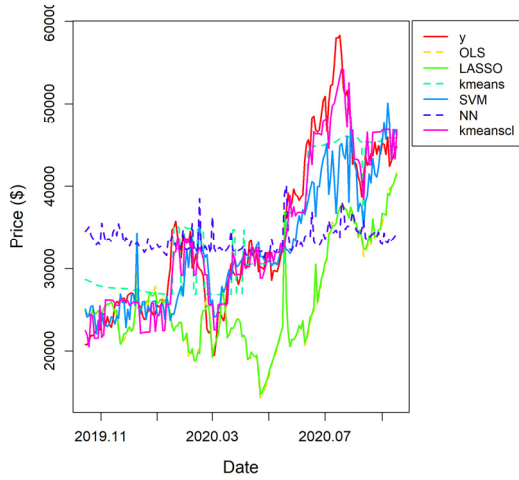
* kmeans(k-means clustering), kmeanscl(Two stage k-means clustering)

Fig. 2는 각 회사별로 2단계 k-평균 군집화 방법과 다른 예측 기법에 의한 예측추세를 비교하고 있다. 3개 회사 모두에서 실제 주식가격(붉은색)에 가장 근접한 추세를 보여주는 기법은 2단계 k-평균 군집화 방법(분홍색)이라는 것을 알 수 있다. 특히, 실제 주식 가격의 큰 변동과 함께 변하는 예측값을 보여주는 것은 2단계 k-평균 군집화 방법으로, 이것은 2단계 k-평균 군집화 방법이 예측 목표값에 가장 근접한 군집에 기반하여 예측을 하는 것이라고 해석할 수 있다. k-평균 군집화 방법(하늘색 색선)의 경우에는 다른 예측기법보다 실제 주식가격에 다소 가까운 모습을 보이거나, 구간별로 나타나는 변동성은 거의 반영하지 못하고 있어 필요 이상으로 큰 군집에 기반한 예측을 수행하여 상대적으로 정확성이 떨어진다 는 것을 보여주고 있다. 이러한 경향은 Y사의 경우에서 잘 드러나는데, 2020년 7월 근방에서 실제 주식 가격이 크게 상승하지만 k-평균 군집화 방법은 과거의 실제 주식 가격에 영향을 받아 주식 가격의 갑작스러운 상승 현상을 반영하지 못하고 있는 것으로 나타났다.

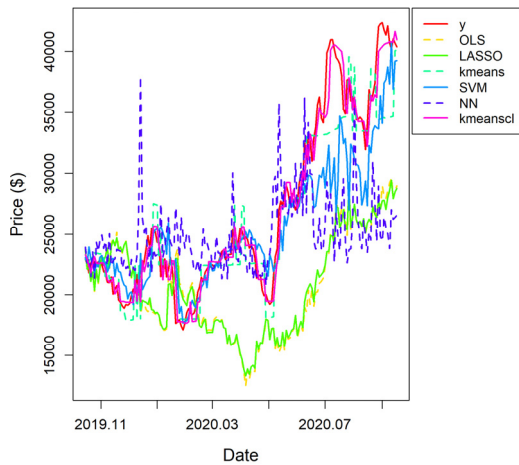
Fig. 3은 상자그림을 통하여 각 회사별로 2단계 k-평균 군집화 방법과 다른 예측 기법에 의한 예측결과를 비교한 것이다. 3개 회사 모두에서 2단계 k-평균 군집화 방법의 증위값이 가장 낮은 것으로 나타나 높은 정확성을 보여주었다. 특히, 2단계의 군집화를 거친 규모가 적은 군집에도 불구하고 2단계 k-평균 군집화 방법은 예측 오차의 분산된 정도가 낮게 측정되어 예측의 정확성과 안전성을 동시에 달성할 수 있는 것으로 나타났다.



(a) Firm S

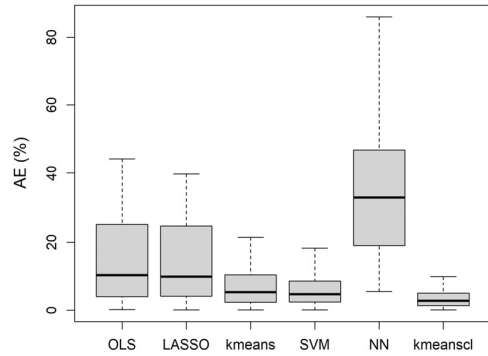


(b) Firm Y

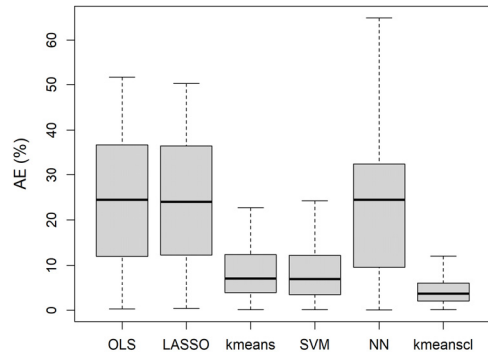


(c) Firm J

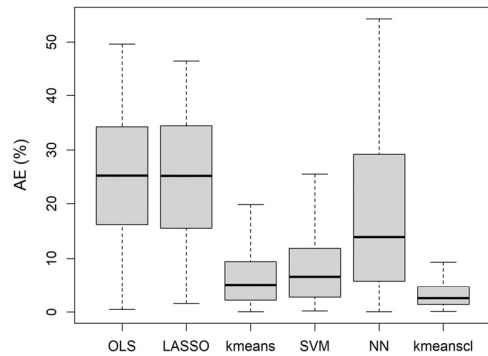
Fig. 2. Comparison of the predicted stock prices



(a) Firm S



(b) Firm Y



(c) Firm J

* AE: Absolte error

Fig. 3. Boxplot comparison

5. 결론

변수들 간의 검증된 인과관계를 바탕으로 얻어진 특정한 함수식을 기반으로 하는 예측기법은 충분한 데이터 확보여부, 연구에서 다루지 못한 외부 변수 등 다양한 요소에 의해 언제나 높은 예측성능을 담보하기는 쉽지 않다. 반면에, 머신러닝은 변수들간의 사전 지식 없이 주어진 데이터의 특성에만 기반한 예측기법이다. 특히, 이러한 머신러닝 기법 중 주어진 데이터에서 서로 유사한 하위 샘플을 군집화하여 분류 및 예측 작업을 수행하는 k-평균 군집화는 널리 쓰이는 머신러닝 기법으로 알려져 있다. 기존 연구들에서 k-평균 군집화의 초기 k값은 임의로 설정되는 경우가 많아 보다 세분화된 하위 군집에 기반한 예측에는 제한적인 부분이 있었다. 본 연구는 기존의 k-평균 군집화에서 얻어진 하위 군집에 다시 k-평균 군집화를 적용하여 예측 대상값과 보다 근접한 세분화된 하위 군집에 기반한 2단계 k-평균 군집화 예측기법을 사용하여 한류콘텐츠 기업의 주식 가격을 예측하였다. 분석결과, 2단계 k-평균 군집화는 다른 예측기법보다 낮은 수준의 예측오차를 보여 예측 정확성이 높은 것으로 나타났다. 특히, 세분화된 하위 군집으로 사용된 자료의 개수가 상대적으로 부족함에도 불구하고 예측오차의 분산된 정도가 낮게 나타나 비교적 안정된 예측성능도 보여주었다.

이러한 점을 고려하면, 2단계 k-평균 군집화는 다음과 같은 이론적인 시사점을 내포하고 있다. 본 기법은 2단계의 군집화를 거치면서 상대적으로 크기가 작은 군집을 사용함에도 불구하고 실제 값에 보다 근접한 군집을 사용함에 따라 예측오차의 분산을 줄일 수 있었다. 일반적으로 샘플의 크기가 적을수록 분산은 증가하는 경향이 있으나, 본 기법은 실제값에 근접한 군집을 사용함으로써 이러한 문제를 완화할 수 있었다. 특히, 이러한 2단계 k-평균 군집화의 장점은 전체 자료를 모두 수집하기 어려운 경우에 적용될 수 있다. 예를 들면, 자료 수집 과정에서 자료를 우선 군집화하고 예측하고자 하는 시점에 근접한 군집만을 추출하여 2단계 k-평균 군집화를 적용하면 전체 자료가 주어지지 않더라도, 예측값을 얻을 수 있는 것이다.

또한, 2단계 k-평균 군집화 방법이 주는 실무적 시사점으로는 한류콘텐츠 기업과 같이 비재무적 정보가 상대적으로 큰 영향을 주는 경우의 주가 예측에서 예측의 정확성과 안정성을 높였다는 점이다. 이에 따라, 기존에는 개인투자자들이 이른바 호재, 악재 등의 비재무적 정보

를 단순하게 해석하여 투자 결정을 하고, 투자손익의 변동성에 크게 노출되었지만, 2단계 k-평균 군집화 방법을 통하여 개인투자자들이 비교적 안정적인 주가 예측을 도모할 수 있는 것이다. 또한, 기존의 머신러닝 기법들의 경우 변수 선택, 조정 파라미터 등 수리적 고려사항이 많았지만 2단계 k-평균 군집화는 기존의 군집화 방법을 반복 적용하여 적절한 군집만 탐색하면 되므로 실무적으로 실용성이 높다고 하겠다.

2단계 k-평균 군집화는 기본적으로 k-평균 군집화를 기반으로 이루어지므로 기존의 연구와 같이 초기의 k값을 적절히 잘 설정하는 것이 중요하다고 볼 수 있다. 이론적으로는 주어진 데이터의 개수만큼 k값을 설정할 수 있으나, 군집내의 자료 개수가 추정 파라미터 대비 적을수록 예측오차의 분산이 커진다는 단점이 있다. 이점을 고려하면, 2단계 k-평균 군집화는 추가적인 군집화 과정을 거치므로 데이터의 개수에 따라 k값의 상한이 기존의 k-평균 군집화보다 더 적게 설정되고, 기존의 k-평균 군집화보다 초기 k값을 더 적은 범위에서 설정할 수 있다는 점에서 알고리즘 효율성을 높인다고도 볼 수 있다. 한편, 2단계 k-평균 군집화는 각 군집화 단계마다 k값의 상한을 동일하게 설정한다면 자료의 개수가 최소 k^2 이 되어야 하므로 소규모 자료가 주어진 연구에는 적용의 한계점이 있다. 아울러, 자료의 크기가 상당히 큰 경우에는 더 높은 단계의 군집화도 가능하므로 자료의 크기와 군집화 반복횟수 간의 분석적 접근도 향후에는 의미있는 연구가 될 수 있을 것으로 보인다.

REFERENCES

- [1] T. Hastie, R. Tibshirani & J. Friedman. (2009). The Elements Of Statistical Learning: Data Mining, Inference, And Prediction. *International Statistical Review*, 77(3), 463-482.
DOI : 10.1111/j.1751-5823.2009.00095_18.x
- [2] S. Y. Kwon, Y. W. Ko & M. H. Hwang. (2010). The Market Reaction to Management Earnings Forecasts and its Determinants. *Korean Management Review*, 39(4), 995-1021.
- [3] J. Y. Heo & J. Y. Yang. (2015). SVM based Stock Price Forecasting Using Financial Statements. *KIISE Transactions on Computing Practices*, 21(3), 167-172.
DOI : 10.5626/KTCP.2015.21.3.167
- [4] J. M. Park, G. H. Kim & N. C. Kang. (2015). The Market Reaction on the Corrective Disclosure of Management Earnings Forecasts. *Korean International Accounting Review*, 64, 183-200.

- [5] J. B. Kim. (2012. 9. 17). Venture start-up fever, light and dark. *Electronic Times News*.
- [6] H. Y. Jung. (2020. 6. 4). Bio-Pharmaceutical Companies, Individuals Still Invest. *ChosunBiz*.
- [7] W. S. Jung. (2020. 4. 20). Young people in their 20s with the highest percentage of "invest in stocks"... Negative bankbook debt growth rate is also 75% "highest". *The Kyunghyang Shinmun*.
- [8] S. S. Kim, D. W. Nam, H. Jo & S. H. Kim. (2012). A study on the relation of web news and stock price. *Journal of Information Technology Service*, 11(3), 191-203.
DOI : 10.9716/KITS.2012.11.3.191.
- [9] D. Y. Kim, J. W. Park & J. H. Choi. (2014). A comparative study between stock price prediction models using sentiment analysis and machine learning based on SNS and news articles. *Journal of Information Technology Service*, 13(3), 221-233.
DOI : 10.9716/KITS.2014.13.3.221.
- [10] S. Shin, H. J. Lee, & J. J. Ahn. (2018). A study on initial price change prediction of IPO shares using non-financial information. *Journal of the Korean Data And Information Science Society*, 29(2), 425-439..
- [11] M. A. Tayal & M. M. Raghuvanshi. (2010). Review on various clustering methods for the image data. *Journal of Emerging Trends in Computing and Information Sciences*, 2, 34-38.
- [12] F. Cai, N. A. Le-Khac & M. T. Kechadi. (2012). Clustering Approaches for Financial Data Analysis. in *Conference: 8th International conference on Data Mining*. Nevada : USA.
- [13] I. C. Park, O. J. Kwon & T. Y. Kim. (2009). KOSPI directivity forecasting by time series model. *Journal of the Korean Data And Information Science Society*, 20(6), 991-998.
- [14] S. Kim & J. A. Kim. (2009). Analyzing financial time series data using the GARCH model. *Journal of the Korean Data And Information Science Society*, 20(3), 475-483.
- [15] H. J. Song & S. J. Lee. (2018). A study on the optimal trading frequency pattern and forecasting timing in real time stock trading using deep learning: focused on KOSDAQ. *The Journal of Information Systems*, 27(3), 123-140.
DOI : 10.5859/KAIS.2018.27.3.123
- [16] D. S. Song. (2002). An empirical study on the effects of accounting information in kosdaq firms on the stock price. *Asia Pacific Journal of Small Business*, 24(4), 79-98.
- [17] A. J. O'connor. (2013). The power of popularity: An empirical study of the relationship between social media fan counts and brand company stock prices. *Social Science Computer Review*, 31(2), 229-235.
- [18] M. J. Kim, J. H. Ryu, D. H. Cha & M. K. Sim. (2020). Stock price prediction using sentiment analysis: from "stock discussion room in naver. *The Journal of Society for e-Business Studies*, 25(4), 61-75.
DOI : 10.7838/JSEBS.2020.25.4.061.
- [19] J. MacQueen. (1967). Some methods for classification and analysis of multivariate observations. in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. (Vol. 1, No. 14, pp. 281-297). Oakland : USA.
- [20] O. Oyelade, O. O. Oladipupo & I. Obagbuwa. (2010). Application of k Means Clustering algorithm for prediction of Students Academic Performance. *International Journal of Computer Science and Information Security*, 7(1), 292-295.
- [21] N. Nidheesh, K. A. Nazeer & P. Ameer. (2017). An enhanced deterministic K-Means clustering algorithm for cancer subtype prediction from gene expression data. *Computers in biology and medicine*, 91, 213-221.
DOI : 10.1016/j.compbiomed.2017.10.014
- [22] R. P. D. Nath, H. J. Lee, N. K. Chowdhury & J. W. Chang. (2010). Modified K-means clustering for travel time prediction based on historical traffic data. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer. (pp. 511-521).
DOI : 10.1007/978-3-642-15387-7_55
- [23] R. Shinde, S. Arjun, P. Patil & J. Waghmare. (2015). An intelligent heart disease prediction system using k-means clustering and Naive Bayes algorithm. *International Journal of Computer Science and Information Technologies*, 6(1), 637-639.
- [24] Y. C. Lee. (2011). Clustering-based Performance Prediction Model Using Technology Rating Data. *Journal of The Korean Data Analysis Society*, 13(3), 1471-1482.
- [25] W. Chen, S. Chen, H. Zhang & T. Wu (2017). A hybrid prediction model for type 2 diabetes using K-means and decision tree. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, IEEE. (pp. 386-390).
DOI : 10.1109/ICSESS.2017.8342938
- [26] A. Jamal, A. Handayani, A. Septiandri, E. Ripmiatin & Y. Effendi. (2018). Dimensionality reduction using pca and k-means clustering for breast cancer prediction. *Lontar Komput. J. Ilm. Teknol. Inf*, 9(3), 192-201.
DOI : 10.24843/LKJITI.2018.v09.i03.p08
- [27] K. Benmouiza & A. Cheknane. (2013). Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. *Energy Conversion and Management*, 75, 561-569.
DOI : 10.1016/j.enconman.2013.07.003
- [28] M. H. Huh. (2000). Double k - means clustering. *The Korean Journal of Applied Statistics*, 13(2), 343-352.
- [29] L. Nanetti, L. Cerliani, V. Gazzola, R. Renken & C.

- Keyser. (2009). Group analyses of connectivity-based cortical parcellation using repeated k-means clustering. *Neuroimage*, 47(4), 1666-1677.
DOI : 10.1016/j.neuroimage.2009.06.014
- [30] H. Ismahan. (2018). I-k-means?: An Iterative Clustering Algorithm Based on an Enhanced Version of the k-kmeans. *Pattern Recognition*, 79, 402.-413.
DOI : 10.1016/j.patcog.2018.02.015
- [31] R. Salman, V. Kecman, Q. Li, R. Strack & E. Test. (2011). Two-Stage Clustering with k-Means Algorithm. *Communications in Computer and Information Science*, 162, 110-122.
DOI : 10.1007/978-3-642-21937-5_11
- [32] N. Chayangkoon & A. Srivihok. (2016). Two Step Clustering Model for K-Means Algorithm. *Proceedings of the Fifth International Conference on Network, Communication and Computing-ICNCC*.
DOI : 10.1145/3033288.3033347
- [33] A. Singh, A. Yadav & A. Rana. (2013). K-means with Three different Distance Metrics. *International Journal of Computer Applications*, 67(10), 13-17.
DOI : 10.5120/11430-6785
- [34] E. Xing, M. Jordan, S. J. Russell & A. Ng. (2002). Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems*, 15, 521-528.
DOI:10.5120/11430-6785
- [35] S. A. Fahad & M. M. Alam. (2016). A modified K-means algorithm for big data clustering. *International Journal of Science, Engineering and Computer Technology*, 6(4), 129-132.
- [36] V. Cherkassky & Y. Ma. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, 17(1), 113-126.
DOI : 10.1016/S0893-6080(03)00169-2
- [37] M. Bank, M. Larch & G. Peter. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Finance Markets and Portfolio Management*, 25(3), 239-264.
DOI : 10.1007/s11408-011-0165-y

김 정 우(Jeong-Woo Kim)

[정회원]



- 2005년 8월 : 고려대학교 경제학(경제학사)
- 2012년 8월 : 연세대학교 경제학(경제학석사)
- 2018년 2월 : 연세대학교 경제학(경제학박사)
- 2020년 3월 ~ 현재 : 강릉원주대학교

경제학과 교수

- 관심분야 : 계량경제학, 머신러닝
- E-Mail : kurtkim@gwnu.ac.kr