Information Systems Review
Vol. 22, No. 1 February 2020 http://dx.doi.org/10.14329/isr.2020.22.1.001

머신러닝 기법을 활용한 블록체인의 엉클블록 분석 연구

A Study on Uncle Block Analysis of Blockchain Using Machine Learning Techniques

김 한 민 (Han-Min Kim) 성균관대학교 경영대학 박사과정

요 약

블록체인은 시스템에 참여하는 사용자들 사이에 믿음을 확보할 수 있는 기술로 떠오르고 있다. 블록체인에 대한 관심이 높아지면서 선행 연구들은 블록체인 기술에 관련된 암호화폐와 적용방안에 대한 연구를 주로 수행하였다. 반면에 블록체인의 안정적인 구동에 대한 연구는 크게 주목하지 않았다. 대표적으로 블록체인의 엉클블록은 블록체인 시스템의 안정적 구동에 중요한 역할을 담당함에도 불구하고 이에 대한 연구는 거의 수행되지 않았다. 이러한 인식을 기반으로 본 연구는 블록체인 정보와 거시 경제 요인들을 활용하여 블록체인의 엉클블록을 머신러닝 기법으로 예측하고자 하였다. 인공신경망, 서포트벡터머신 분석 결과, 블록체인 정보와 거시 경제 요인들은 블록체인의 엉클블록 예측에 기여하는 것으로 나타났다. 또한, 블록체인 정보만을 활용한 인공신경망은 엉클블록의 발생을 예측하는데 가장 우수한 성능을 제공하는 것으로 나타났다. 본 연구는 정보시스템 분야에서 블록체인 연구를 주도하고 기여할 수 있는 방안을 제시한다.

키워드 : 블록체인, 엉클블록, 머신러닝, 예측, 블록체인정보

Ⅰ. 서 론

블록체인이 비즈니스 참여자들의 신뢰 문제를 해결할 수 있는 기술로 주목받으면서 다양한 분야에서 블록체인에 대한 관심이 크게 증가하고 있다. 블록체인은 분산화 된 원장을 네트워크 참여자에게 공유하고 검증하는 기술로써 블록체인이 등장하기 이전에는 특정 네트워크의 참여자들이 일관화 된 원장을 공유하고 서로를 신뢰하게 만드는 문제에 대해서 큰 어려움을 겪고 있었다(Abraham *et al.*, 2018; Antonopoulos, 2014). 블록체인이 사회적으로 주목받기 시작한 것은 블록체인 시스템에 존재하는 코인 시스템이 주목 받으면서부터였다. 최초의 1세대 블록체인 코인인 비트코인이 생겨나고 2세대 블록체인이라고 할 수 있는 이더리움이 등장하면서 많은 유사 코인들이 파생되었다(Antonopoulos and Wood, 2018; Ciaian *et al.*, 2016). 다양한 코인들이 생겨나면서 코인에 대한 관심이 증가하였고 자연스럽게 블록체인 기술의 본질적 가치도 주목받기 시작하였다. 블록체인 관련 선행연구들은 비트코인 및 파생 코인들의 가격을 예측하는 연구들을 수행하거나 블록체인이 산업에 적용될 수 있는 방

안에 대한 연구를 주로 진행하였다(Abraham *et al.*, 2018; Ciaian *et al.*, 2016; Jang and Lee, 2017; Kim *et al.*, 2016; Li *et al.*, 2018).

블록체인은 블록에 기록된 정보가 조작이 거의 불가능하기 때문에 네트워크 참여자들이 신뢰할 수 있는 정보를 공유할 수 있다는 장점을 가지고 있지만 시스템적으로 보완되어야 할 부분이 있다. 그 중 엉클블록 문제는 블록체인 시스템이 관리해야 할 대표적인 문제 중 하나이다(Antonopoulos, 2014). 블록체인에서 정보가 정식으로 등록되기 위해서는 시스템 내에서 새로운 블록을 만들어 네트워크에 전파시키고 정식 블록으로 인정받아야 한다. 블록체인 네트워크에 정식으로 등록되지 못한 블록들은 자연스럽게 엉클블록으로 남게 된다. 엉클블록은 정식노드에 등록되지 않기 때문에 엉클블록에 저장된 정보들은 다른 새로운 블록에 저장되어 정식 노드로 인정받을 때까지 불완전한 상태로 남아있게 된다. 이러한 경우 거래가 이루어짐에도 불구하고 거래 정보가 다시 롤백(Roll back) 되는 결과를 초래할 수 있다. 신속하게 변화하는 비즈니스 환경에서 이러한 롤백 현상은 블록체인을 활용하는데 큰 걸림돌이 될 수 있다(Antonopoulos and Wood, 2018). 엉클블록의 문제가 과도하게 심화되면 하드 포크(Hard fork) 현상이 발생하게 된다. 하드 포크란 정식 노드 이외에 엉클블록의 체인화가 심화되어 다른 노드가 생겨나는 것을 말한다(Antonopoulos and Wood, 2018). 실제로 비트코인과 이더리움은 하드 포크의 결과로 블록체인 시스템 내에 다른 노드들이 생겨난 경우가 존재한다. 이러한 하드 포크는 블록체인의 가치와 안정성을 저하시키는 요인으로 현재까지 실제 하드포크의 사례는 종종 보고되고 있다. 예를 들어 이더리움 클래식의 경우 최초의 블록과 초기 블록체인의 정보는 이더리움과 동일하지만 하드포크가 발생한 지점부터는 정식 이더리움 노드와 완전히 다른 노드가 생성되어 전혀 다른 블록체인으로 구동되고 있다(Breitman, 2017).

이러한 중요성에도 불구하고 아직까지 블록체인의 엉클블록에 대한 연구는 거의 이루어지지 않은 상황이다. 따라서 본 연구는 블록체인의 엉클블록을 예측하고 관리할 수 있도록 엉클블록의 발생을 머신러닝 기법으로 예측하고자 한다. 본 연구는 여러 산업분야서 활용되고 주목하고 있는 블록체인인 이더리움을 대상으로 연구를 진행하고자 한다. 이더리움은 비트코인과 다르게 블록에 다양한 종류의 정보를 담을 수 있는 확장성이 존재한다. 때문에 실용성이 높은 블록체인으로 주목받고 있다(Antonopoulos and Wood, 2018). 본 연구는 이더리움의 엉클블록 발생 예측을 위한 방법으로 이더리움의 블록체인 정보와 거시 경제 요인을 고려하고 인공 신경망(artificial neural network, ANN)과 서포트 벡터 머신(support vector machine, SVM)의 머신러닝 기법을 활용하고자 한다.

본 연구의 연구 질문은 다음과 같다. 첫 번째, 이더리움 엉클블록 발생 예측에 이더리움 블록체인 정보가 관련될 수 있는가? 두 번째, 이더리움 엉클블록 발생 예측에 거시 경제 요인들이 관련될 수 있는가? 세 번째, 엉클블록 발생을 가장 잘 예측할 수 있는 머신러닝 기법 및 요인들은 무엇인가? 본 연구의 결과를 토대로 학문적 실무적 기여사항을 논의하고 한계점 및 연구방향을 제시하고자 한다. 본 연구의 발견은 블록체인의 엉클블록에 대한 연구의 필요성을 제시하는 탐험적이고 선행적인 연구가 될 것이다. 학문적으로는 엉클블록과 관련되는 변수들을 구체적으로 제시하고 적합한 머신러닝 기법을 제시할 수 있을 것이다. 실무적으로는 엉클블록의 발생을 예측하고 사전에 대비하여 블록체인의 안정적인 구동과 실무적용에 기여할 수 있을 것이다.

본 논문의 구성은 다음과 같다. 제Ⅱ장에서 블록체인에 대한 이론적 설명과 선행연구를 살펴보고자 한다. 제Ⅲ장에서는 분석을 위한 방법론을 서술한다. 제Ⅳ장에서는 분석 결과를 제시하며, 제Ⅴ장에서 본 연구의 기여사항과 한계점 및 향후 연구방향에 대해 제언한다.

Ⅱ. 이론적 배경

2.1 블록체인

블록체인은 모든 참여자가 분산 네트워크를 통해 공유된 정보와 가치를 검증함으로써 임의적인 조작이 어렵도록 설계된 분산 시스템 기술이다(Abraham *et al.*, 2018). 블록체인은 은행과 같은 기존의 중앙 집중식 거래방식과 다르게 모든 참여자들이 조작이 거의 불가능한 거래 원장을 공유할 수 있다. 이러한 특성 때문에 블록체인은 중개자 없이 거래의 신뢰성을 확보할 수 있다.

블록체인을 구성하는 블록은 블록 헤더와 바디로 구성된다(Antonopoulos, 2014). 블록의 헤더 부분은 블록 생성의 난이도, 넌스 값, 이전 블록의 해시 값, 현재 블록의 생성 시간, 거래 정보들을 해싱 한 머클루트 값이 포함되며, 블록 바디는 거래정보를 포함한다. 블록 헤더의 정보를 이용해 현재 블록의 고유한 해시 값을 도출해낼 수 있으며, 이러한 해시 값은 해시 함수를 적용한 값이다. 반복 계산을 통해 목표 값을 충족하는 넌스 값을 찾아내면 새로운 블록이 생성이 된다. 새롭게 생성한 블록을 네트워크 참여자들 절반 이상이 검증하여 네트워크에 전파하게 되면 정식블록으로 인정받게 된다. 채굴자는 이에 대한 보상을 코인으로 지급받게 되며 이러한 과정을 작업증명(Proof of Work) 라고 한다.

블록체인에서 사용되는 해시 함수는 입력 값에 아주 미세한 변화가 발생할 경우 결과 값이 완전히 다르게 도출되기 때문에 결과 값을 보고 일정한 패턴이나 공식을 이용하여 입력 값을 찾아내는 것은 거의 불가능하게 구성되어 있다(Antonopoulos, 2014). 해시 함수는 입력 값을 알고 있을 경우 결과 값을 계산하는 것은 매우 간단하지만 결과 값만을 가지고 입력 값을 찾아내는 것은 불가능에 가깝다. 결과 값으로 입력 값을 찾아낼 수 있는 방법은 임의의 입력 값을 반복적으로 대입 계산하여 올바른 결과 값이 나오는 입력 값을 발견하는 방법뿐이다. 따라서 블록의 거래정보를 조작할 경우 결과 값은 블록의 완전히 다른 해시 값으로 도출되고 원래의 블록 해시 값으로 설정하기 위해서는 엄청난 컴퓨터 계산 시도가 필요 하다. 블록체인은 새로운 블록이 생성될 때마다 이전 블록의 고유 해시 값을 포함하기 하기 때문에 블록 간의 체인 구조가 형성된다. 체인 구조로 형성된 블록 내의 정보를 조작하기 위해서는 블록체인에 등록되어 있는 모든 블록을 조작하여야 한다. 또한 새로운 블록이 지속적으로 생성되어 체인이 형성되고 있기 때문에 블록체인에 기록된 정보를 조작하는 것은 현재 기술로는 거의 불가능하다고 할 수 있다(Antonopoulos and Wood, 2018).

하지만 블록체인도 몇 가지 단점을 가지고 있다(Antonopoulos and Wood, 2018; Jang and Lee, 2017). 첫 번째, 공격자가 블록체인 네트워크에 참여하는 참여자들의 컴퓨터 계산 능력보다 51% 이상의 컴퓨터 계산 능력을 보유하고 있다면 블록체인에 기록되는 정보들은 조작될 가능성이 존재한다. 두 번째, 현재 비즈니스에서 요구하는 정보의 빠른 교환과 처리에 있어 블록체인의 정보 처리 속도는 충분치 않다. 비트코인은 평균 10분마다 이더리움은 평균 15초마다 블록을 생성하고 있기 때문에 신속한 거래의 처리를 위해서 해결되어야 할 문제로 남아 있다. 마지막으로 엉클블록의 문제는 새롭게 채굴된 블록이 정식 노드로 인정받지 못할 때 발생한다. 엉클블록에 기록된 거래정보는 정식 노드에 기록된 것이 아니기 때문에 발생한 거래기록이 엉클블록에 저장된 것이라면 거래가 롤백 되거나 새로운 블록에 다시 등록되어 정식 노드로 인정받을 때가지 시간이 필요하다. 엉클블록 문제가 극도로 심화되면 하드포크 문제가 발생한다. 하드포크 문제는 하나의 블록체인 시스템에 정식 노드 이외의 전혀 다른 노드가 추가로 발생하는 것으로 퍼블릭 블록체인의 가치와 안정성의 저하를 야기 시킬 수 있다(Antonopoulos and Wood, 2018; Breitman, 2017). 하지만 이러한 중요성에도 불구

하고 아직까지 엉클블록에 대한 연구는 거의 수행되지 않은 상황이다. 블록체인이 여러 분야에서 중대한 영향을 미치는 정보기술인 만큼 정보시스템 분야에서는 블록체인에 대한 연구가 활발하고 지속적으로 수행될 필요가 있다. 본 연구는 블록체인의 엉클블록 발생 문제에 대한 탐색적 연구로써 정보시스템 분야에서 블록체인 연구에 기여할 수 있는 방안을 제시하고자 한다.

2.2 블록체인 관련 연구

선행 연구는 주로 블록체인 시스템의 한 부분인 코인 가격 예측을 실증하거나 산업에 적용할 수 있는 가능성과 방안을 주로 수행하였다(Abraham *et al.*, 2018; Cianian *et al.*, 2016; Jang and Lee, 2017; Mensi *et al.*, 2019). 코인 가격 예측 연구에서는 주로 비트코인, 이더리움, 모네로의 코인을 대상으로 분석을 수행하였다. 그 결과, 소셜미디어의 댓글 및 활동이 코인 가격과 관련되어 있다는 사실을 발견하였으며(Abraham *et al.*, 2017; Kim *et al.*, 2016; Kristoufek, 2013), 코인이 블록체인 시스템 내에서 발행되고 소모되는 점을 반영하여 블록체인 정보가 코인 가격에 관련된다는 사실을 발견하였다(Jang and Lee, 2017). 또한, S&P 500, DOW30, Eurostoxx, 금 가격 등의 거시경제 요인들도 코인가격 예측에 관련되어 있다는 사실을 발견하였다(Ciaian *et al.*, 2016; Jang and Lee, 2017).

한편, 블록체인이 산업 분야에 적용될 수 있는 방안과 연구 방향을 제시하는 연구들도 이루어졌다. 일부 연구에서는 헬스 케어 산업에서 블록체인이 적용될 수 있는 사례를 제시하면서 블록체인을 활용하여 의료데이터의 관리 및 권한을 사용자에게 돌려줄 수 있는 방안을 제시하였다(Kuo *et al.*, 2017; Yue *et al.*, 2016). 또 다른 연구에서는 현재 블록체인에 대한 문제와 향후 연구방향을 파악하기 위해 개인을 대상으로 블록체인에 대한 인식과 극복해야할 문제들에 대해 설문조사를 실시하였다(Zheng *et al.*, 2018). IoT(Internet of Things) 관련 연구에서는 IoT에 활용될 수 있는 블록체인의 적용방안을 검토하고 보안 문제를 개선하고자 하는 연구도 진행되었다(Li *et al.*, 2018). 공급망 관리(Supply Chain Management) 측면 에서는 블록체인의 장점을 활용하여 투명하고 안전한 공급망 관리 시스템 구축에 대한 방안을 제시하였다(Wu *et al.*, 2017).

하지만 아직까지 블록체인 시스템의 엉클블록 발생에 관련된 요인들을 발견하고 분석하고자 하는 연구는 수행되지 않았다. 앞 장에서 언급한 것처럼 블록체인 시스템의 엉클블록 문제는 향후 블록체인 산업이 발전하고 지속적으로 성장하기 위해 해결 되어야할 과제로 남아있다. 따라서 본 연구는 머신러닝 기법을 통해 블록체인의 엉클블록 발생에 관련되는 변수들을 발견하고 예측하고자 한다. 본 연구는 연구 목적을 달성하기 위하여 예측 문제에서 주로 적용되고 뛰어난 예측 분석 결과를 나타낸 인공신경망과 서포트 벡터 머신 분석기법을 활용하고자 한다.

2.3 머신러닝과 인공신경망

예측 관련 문제를 분석하기 위해서 선행 연구들은 전통적인 회귀 모델이나 자기 회귀 이동평균(Autoregressive Integrated Moving Average: ARIMA) 모델 방법 이외에도 머신러닝 기법을 적극 활용하였다(Jang and Lee, 2017; Kaytez *et al.*, 2015; Pati *et al.*, 2017). 그 결과, 전통적인 회귀 모델 분석보다 머신러닝 분석 기법을 통해 향상된 분석 결과를 얻을 수 있다는 사실을 확인해왔다(Kaytez *et al.*, 2015; Pati *et al.*, 2017; Wang, 2011; Yao and Tan, 2000). 그 중 인공신경망(Artificial Neural Networks: ANN)과 서포트 벡터 머신(Support Vector Machine: SVM)은 머신러닝을 활용한 예측 관련 분석에서 대표적으로 활용되어 뛰어난 예측력을 제공하였다.

인간의 뇌신경 구조를 모방하여 만든 인공 신경망은 입력 층(Input layer)과 출력 층(Output layer)

사이에 한 개 또는 다수의 은닉 층(Hidden layer)이 존재하는 다층 퍼셉트론(Multilayer Perceptron: MLP) 구조로 되어 있다. 입력 층에 입력된 값은 은닉 층을 통해 출력 층의 결과 값으로 계산된다. 결과 값이 실제 값과 차이가 크게 발생하게 되면 인공 신경망은 에러가 큰 결과 값을 입력 층으로 다시 전파시킨다. 그 후 입력 값에 다시 가중치를 조정해서 계산을 수행한다. 이러한 과정을 역전파 과정이라고 지칭하며, 이를 통해서 인공 신경망은 결과 값과 실제 값 사이의 오차를 최소화하는 학습을 수행하고, 결과적으로 오차가 최소화된 결과 값을 도출하게 된다. 인공 신경망은 다수의 은닉층을 가질 수 있으며 이러한 구조를 통해 단순 선형 문제를 넘어서 자연어 처리, 재무 분석, 이미지 인식 등 다양한 비선형 문제 처리할 수 있게 된다 (Kara *et al.*, 2011; Murphy, 2012).

선행 연구에서는 인공신경망을 활용하여 일반적인 경제 지표를 예측하거나 코인가격 예측하는 등 다양한 문제를 해결해왔다. Yao and Tan(2000)은 Nikkei 225 지표의 미래 옵션을 예측하기 위해 인공신경망과 블랙 숄즈 모델(Black-Scholes Model)을 비교 분석하였다. 그 결과 인공 신경망이 블랙 숄즈 모델보다 뛰어난 예측력을 제공한다는 사실을 발견하였다. Pati *et al.*(2017)는 소프트웨어의 복제 예측에 자기 회귀 이동평균 모델과 인공 신경망을 비교 분석한 결과 인공 신경망이 보다 향상된 예측력을 제공한다는 사실을 발견하였다. 블록체인 관련 연구에서는 비트코인의 블록체인 정보, 거시경제 요인 데이터를 활용한 인공 신경망이 비트코인 가격을 가장 정확하게 예측하는 것으로 나타났다(Jang and Lee, 2017).

2.4 서포트벡터 머신

서포트 벡터 머신은 분류와 회귀분석을 수행할 수 있는 머신 러닝 기법으로 선형 및 비선형 문제 해결에 적합한 분석 방법이다(Kara *et al.*, 2011; Park and Hong, 2009). Cortes and Vapnik(1995)은 분류 문제를 해결하고자 초평면(hyperplane)의 원리를 사용한 서포트 벡터 머신을 제시하였다. 서포트 벡터 머신의 기본적인 원리는 분류될 수 있는 데이터들 간의 거리가 최대화되는 결정면(margin)을 찾아 초평면을 생성하는 것이다. 데이터가 선형으로 분리되지 않는 경우 서포트 벡터 머신은 커널(Kernels)을 활용하여 데이터의 차원을 높이게 된다. 높아진 차원에서 서포트 벡터 머신은 데이터를 분리할 수 있는 초평면을 찾아낸다. 서포트 벡터 머신에서 주로 사용되어지는 커널은 다항식(Polynomial)과 방사 기저(Radial basis)로 나뉘어진다. 아래 식 (1)에서 d는 차수(Degree)이며, 식 (2)에서 σ는 감마 모수(Gamma parameters)이다. 다항식 커널은 식 (1)에 표현된 것처럼 다항식의 차수로 표현되는 d가 식별되어야 한다. 반면에 방사 기저 기반 커널은 식 (2)처럼 표준편차인 σ가 식별되어야 한다.

$$\text{Polynomial} = (x_i \cdot x_j + 1)^d \tag{1}$$

$$Gaussian = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{2}$$

서포트 벡터 머신은 다양한 분야의 연구에서 뛰어난 예측 결과를 제공해왔다. Wang(2011)은 서포트 벡터 머신이 GBP/USD, USD/JPN, AUD/USD, EUR/USD의 환율을 예측하는데 뛰어난 예측력을 제공한다는 사실을 발견하였다. Kaytez *et al.*(2015)은 전기 에너지 소비를 예측하기 위해 서포트 벡터 머신, 인공 신경망, 회귀 분석을 활용하였으며, 연구결과 서포트 벡터 머신이 가장 우수한 예측 성능을 보인다는 것을 발견하였다. 블록체인 관련 연구에서는 비트코인 가격 예측을 위해 서포트 벡터머신과 인공 신경망을 활용하였으며, 결과적으로 서포트 벡터머신이 비트코인 가격을 예측하는데 가장 우수한 분석 방법이라는 사실을 발견하였다 (Mallqui and Fernandes, 2019). 앞서 살펴본 바와 같이 인공 신경망과 서포트 벡터 머신은 다양한 예측 문제와 블록체인 코인 가격 예측 연구에 적용

되어 왔다. 하지만 아직까지 블록체인의 엉클블록 발생 예측에 대해서는 활용되고 있지 않다. 따라서, 본 연구는 대표적인 두 예측 분석 기법을 도입하여 엉클블록 발생예측 분석을 실시하고 적합한 분석 알고리즘을 발견하고 한다.

Ⅲ. 연구 방법

본 연구는 머신러닝 분석을 기반으로 거시 경제 요인, 블록체인 정보를 활용하여 블록체인의 엉클블록 발생을 예측하고자 한다. 분석 프로그램은 래피드 마이너(Rapid Miner) 9.2.001 버전과 SPSS 18.0 버전을 활용하였으며, 머신러닝 분석 방법으로는 선행 연구를 참고하여 인공 신경망과 서포트 벡터 머신을 적용하였다(Jang and Lee, 2017; Mallqui and Fernandes, 2019).

분석에 활용된 데이터는 2015년도 7월 30일부터 2018년 11월 28일까지의 이더리움 블록체인 정보와 거시 경제 요인 데이터이다. 분석 단위는 변수들 모두 요일 단위이며, 이더리움 블록체인 정보 데이터는 https://etherscan.io/ 웹 사이트에서 확보하였다. 거시 경제 요인 데이터는 대학교 학술 정보관의 Data Stream 데이터 베이스에서 확보하였다. 종속변수는 요일 별 이더리움 엉클블록의 발생 수이며 예측 분석을 위하여 log 함수를 적용하였다. 거시 경제 요인의 경우 데이터의 특성 상 일부 누락된 결측치가 존재하기 때문에 결측치 전후의 데이터 값을 평균화하여 대체하였다. 이를 통해 데이터의 결측치가 발생하는 것을 방지하였다. 블록체인의 정보의 경우 거시 경제 요인 데이터와 다르게 주말이나 공휴일에도 데이터가 발생되고 기록된다. 따라서, 블록체인의 정보는 누락된 결측치가 존재하지 않았다. 본 연구는 시계열 분석을 위해 종속변수의 데이터 날짜(t)를 기준으로 설정하였으며, 독립변수들은 종속변수의 하루 전 날짜의 데이터(t-1)로 설정하였다. 독립변수로 활용된 블록체인 정보와 거시 경제 요인들을 아래 <표 1>에 정리하였다.

분석에 앞서 본 연구는 데이터들의 특성과 변수의 정의를 명확히 하고자 한다. 따라서, 아래 <표 2>에 블록체인 정보와 거시 경제 요인들의 정의와 기술적 통계(평균, 표준편차)를 분석하여 기술하였다. 추가적으로, 데이터들 간의 관계를 파악하기 위해 변수들 간 피어슨(Pearson) 상관관계 분석을 실시하였다. 이를 통해 변수들 간의 관계성을 보다 구체적으로 파악하고 분석 결과의 해석 및 결론 도출에 기여할 수 있을 것이다. <표 3>에 변수들 간 상관관계를 기술하였다. <표 3>을 살펴보면, 생성된 블록 수와 관련되는 블록 생성 난이도, 블록의 데이터 크기, 거래 수수료, VIX, Nikkei 225의 상관관계는 유의 수준 0.05에서 유의하지 못한 것으로 나타났다. 또한, 생성된 코인 양과 VIX의 상관관계도 유의하지 않았다. 반면에 다른 모든 변수들 간의 상관관계는 통계적으로 유의한 것으로 나타났다. 그 중에서 이더리움 코인 가격과 엉클블록의 수(0.786), 블록체인 활동 계정 수 와 엉클블록의 수(0.859)는 강한 양의 상관관계를 보이고 있다. 이러한 결과는 본 연구의 엉클블록 발생 예측 분석에서 이더리움 가격과 블록체인 활동 계정 수가 중요한 역할을 담당할 수 있다는 가능성을 나타낸다. 선행연구에서 제시한 결과와 동일하게 이더리움 가격과 거시 경제 요인들 또한 상호 간에 유의한 상관관계를 가지고 있었다. 또한, 이더리움의 가격은 생

〈표 1〉 이더리움 블록체인 정보 및 거시 경제 요인 요약

독립변수(단위: 일)	변수 종류
이더리움 블록체인 정보	블록체인 내 거래 금액, 성사된 거래 수, 생성된 블록 수, 블록 생성 난이도, 블록의 데이터 크기, 생성된 코인 양, 거래 수수료, 이더리움 가격(USD), 블록체인 활동 계정 수
거시 경제 요인	S&P 500, DOW 30, Eurostoxx 50, NASDAQ, Crude oil, SSE, Gold, VIX, Nikkei 225, FTSE100

성된 블록 수를 제외한 블록체인의 정보들과도 긴밀한 관련이 있는 것으로 나타났다. 따라서, 본 연구는 엉클블록 예측 분석에 있어 독립 변수들의 적용을 보다 세심하고 체계적으로 수행할 필요가 있다고 판단한다.

본 연구는 이더리움 블록체인 정보와 거시경제 요인의 영향력을 구체적으로 발견하기 위해서 분석 모델을 세 파트로 구분하여 분석을 실시하였다. 첫 번째 파트에서는 거시 경제 요인들만을 위계적으로 고려하여 엉클블록 예측 분석을 진행하고자 한다. 두 번째 파트에서는 블록체인 정보만을 위계적으로 적용하여 엉클블록 예측분석을 진행하고자 한다. 마지막 세 번째 파트에서는 거시경제 요인과 블록체인 정보를 상호 보완적으로 적용하여 분석을 진행할 예정이다. 이러한 과정을 통해 엉클블록 예측에 핵심적으로 관련되는 변수들을 파악하고 명확한 결과를 제시할 수 있을 것이다.

본 연구는 결과의 신뢰성을 확보하기 위해 10겹 교차 검증(10-Fold Cross Validation)을 적용하였다. 학습 데이터와 테스트 데이터의 파트가 중복되지 않게 10개의 데이터 셋을 형성한 후 10번의 교차 검증을 수행하였다. 이러한 교차 검증을 통해 학습 데이터와 테스트 데이터의 설정 편향에 의한 분석 결과의 문제를 극복하고 신뢰할 수 있는 결과를 얻을 수 있다(김진화 등, 2008).

〈표 2〉 데이터의 정의 및 기술적 통계 요약

변수(단위: 일)	정의	평균(표준편차)
엉클 블록	요일 별 정식 블록체인 네트워크에 등록되지 못한 블록의 수	656.620(398.209)
이더리움 가격(USD)	요일 별 이더리움 가격	210.728(275.677)
블록체인 내 거래 금액(USD)	요일 별 블록체인에서 성사된 거래의 금액	1601085485.654 (3115251763.810)
성사된 거래 수	요일 별 블록체인으로 성사된 거래 수	287817.39(316946.884)
생성된 코인 양	요일 별 발행된 이더리움 코인 양	25841.756(5174.009)
거래 수수료	요일 별 블록에 정보를 등록하기 위해 지불해야 하는 이더리움 수수료 금액	320.383(526.060)
블록체인 활동 계정 수	요일 별 이더리움 블록체인에서 활동하는 계정 수	133696.920(155329.024)
블록의 데이터 크기	요일 별 생성된 블록의 평균 메모리 크기	9663.380(9792.846)
생성된 블록 수	요일 별 생성된 블록의 수	5575.610(707.116)
블록 생성 난이도	요일 별 평균 블록 생성 난이도	1187.409(1362.686)
S&P 500	요일 별 스탠다드 앤 푸어스 500(Standard & Poor's 500) 평균 주가 지수	2373.522(300.692)
DOW 30	요일 별 다우존스(Dow Jones) 산업평균지수	20942.509(3251.418)
Eurostoxx 50	요일 별 유럽 50개 우량 기업 주가지수	3303.187(234.766)
NASDAQ	요일 별 나스닥(NASDAQ) 시장 종합주가지수	72.109(11.520)
Crude oil	요일 별 원유 가격	51.834(10.848)
SSE	요일 별 상하이 증권 거래소의 주식 시장 지수	1418.298(121.1094)
Gold	요일 별 금 가격	1240.305(74.390)
VIX	요일 별VIX(Volatility Index) 시장 기대 지수	14.833(4.686)
Nikkei 225	요일 별 니케이 225(Nikkei225) 주가지수	19678.729(2343.785)
FTSE 100	요일 별 파이낸셜 타임스 주식 거래 100사(Financial Times Stock Exchange 100) 주가 지수	6976.408(564.085)

〈표 3〉 상관관계 분석 결과

	엉클블록 수	생성된 블록 수	블록 생성 난이도	블록의 데이터 크기	생성된 코인 양	성사된 거래 수	거래 수수료	이더리움 가격	블록체인 활동 계정 수	블록체인 내 거래 금액
엉클블록 수	1									
생성된 블록 수	.342**	1								
블록 생성 난이도	.690**	-.010	1							
블록의 데이터 크기	.785**	-.013	.933**	1						
생성된 코인 양	-.498**	.354**	-.855**	-.849**	1					
성사된 거래 수	.853**	.113**	.871**	.964**	-.782**	1				
거래 수수료	.653**	.007	.555**	.675**	-.521**	.696**	1			
이더리움 가격	.786**	.014	.745**	.869**	-.698**	.924**	.696**	1		
블록체인 활동 계정 수	.859**	.103**	.852**	.952**	-.756**	.993**	.704**	.923**	1	
블록체인 내 거래 금액	.411**	-.178**	.292**	.540**	-.424**	.602**	.589**	.689**	.617**	1
S&P 500	.693**	.165**	.910**	.910**	-.724**	.879**	.591**	.759**	.859**	.419**
DOW 30	.701**	.166**	.912**	.920**	-.743**	.898**	.598**	.780**	.876**	.443**
Eurostoxx 50	.349**	-.226**	.506**	.579**	-.591**	.564**	.416**	.586**	.546**	.498**
NASDAQ	.704**	.230**	.902**	.856**	-.633**	.817**	.528**	.662**	.801**	.251**
Crude oil	.733**	.330**	.843**	.811**	-.585**	.796**	.527**	.664**	.784**	.237**
SSE	-.549**	-.096**	-.810**	-.768**	.767**	-.738**	-.438**	-.585**	-.715**	-.259**
VIX	-.069*	-.014	-.064*	-.170**	-.035	-.133**	-.175**	-.139**	-.128**	-.296**
Nikkei 225	.671**	.051	.823**	.860**	-.793**	.848**	.583**	.747**	.829**	.461**
FTSE 100	.514**	.079**	.665**	.713**	-.491**	.691**	.504**	.640**	.676**	.464**
Gold	.339**	.162**	.361**	.412**	-.163**	.439**	.272**	.494**	.428**	.355**

	S&P 500	DOW 30	Eurostoxx 50	NASDAQ	Crude oil	SSE	VIX	Nikkei 225	FTSE 100	Gold
S&P 500	1									
DOW 30	.996**	1								
Eurostoxx 50	.649**	.642**	1							
NASDAQ	.922**	.914**	.414**	1						
Crude oil	.893**	.884**	.494**	.850**	1					
SSE	-.742**	-.769**	-.313**	-.680**	-.634**	1				
VIX	-.335**	-.299**	-.462**	-.275**	-.195**	-.201**	1			
Nikkei 225	.894**	.901**	.795**	.730**	.802**	-.724**	-.226**	1		
FTSE 100	.859**	.841**	.748**	.762**	.739**	-.368**	-.599**	.748**	1	
Gold	.416**	.419**	.071*	.495**	.359**	-.133**	-.286**	.099**	.480**	1

*p > 0.05, **p > 0.01.

본 연구는 분석 결과를 측정하기 위해 RMSE (Root Mean Squared Error)와 MAPE(Mean Absolute Percentage Error)의 측정기준을 활용하였다. 측정기준에 대한 수식은 아래 다음과 같이 표기된다. N은 샘플의 수이며, y_i는 실제 값, $\hat{y}_i$은 추정된 값이다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad (3)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (4)$$

Ⅳ. 분석 결과

본 연구는 앞서 언급했던 바와 같이 거시 경제 요인들만을 위계적으로 고려한 첫 번째 파트 분석과 블록체인 정보만을 위계적으로 고려한 두 번째 파트 분석, 마지막으로 거시 경제 요인들과 블록체인 정보들을 상호적으로 고려한 세 번째 파트로 분석결과가 구성되어 있다. √ 표기는 해당 변수를 분석 모델에 투입한 것을 의미한다. 먼저 첫 번째 파트 분석인 거시 경제 요인들만을 고려한 분석에서는 SVM 보다 ANN 분석 기법이 보다 좋은 분석 결과를 제시하는 것으로 나타났다. 모델 1-1부터 모델 1-10까지의 위계적 분석을 통해 S&P 500, DOW 30, Eurostoxx 50, NASDAQ 변수들은 엉클블록 발생과 관련이 있는 것으로 나타났다. 하지만 거시 경제 요인이 추가될수록 분석 결과는 크게 개선되지 않았다. 따라서, 블록체인의 엉클블록 발생이 거시경제 요인과 긴밀하게 관련되어 있다고 주장하기에는 한계가 존재한다. 거시경제 요인을 활용한 엉클블록 예측 결과를 아래 <표 4>에 정리하였다.

두 번째 파트 분석에서는 블록체인 정보만을 활용하여 엉클블록의 발생을 예측하고자 하였다. 아래 <표 5>에서 모델 2-1부터 모델 2-9까지는 블록체인 정보 변수들을 위계적으로 추가하여 분석한 것이다. 분석 결과, ANN 은 SVM 보다 분석 모델 전반에서 좋은 결과를 보이고 있다.

〈표 4〉 거시경제 요인 활용 분석 결과

		모델 1-1	모델 1-2	모델 1-3	모델 1-4	모델 1-5	모델 1-6	모델 1-7	모델 1-8	모델 1-9	모델 1-10
S&P 500		√	√	√	√	√	√	√	√	√	√
DOW 30			√	√	√	√	√	√	√	√	√
Eurostoxx 50				√	√	√	√	√	√	√	√
NASDAQ					√	√	√	√	√	√	√
Crude oil						√	√	√	√	√	√
SSE							√	√	√	√	√
Gold								√	√	√	√
VIX									√	√	√
Nikkei 225										√	√
FTSE100											√
ANN	RMSE	0.127	0.117	0.111	0.103	0.106	0.097	0.090	0.088	0.091	0.087
	MAPE	0.034	0.031	0.029	0.011	0.011	0.025	0.023	0.023	0.024	0.023
SVM	RMSE	0.175	0.174	0.174	0.167	0.178	0.158	0.160	0.159	0.142	0.143
	MAPE	0.049	0.049	0.049	0.047	0.048	0.044	0.043	0.043	0.039	0.035

블록의 데이터 크기와 거래 수수료는 엉클블록 발생 예측 결과를 향상시키지 못하는 것으로 나타났다. 반면에 다른 블록체인의 정보들은 엉클블록 발생 예측에 기여하는 것을 발견할 수 있었다. 그 중에서 블록체인에서 성사된 거래 수, 생성된 코인, 이더리움 가격은 엉클블록 발생과 중요하게 관련되어 분석 모델의 향상에 크게 기여하는 것으로 나타났다. 추가적으로, 모델 2-9의 예측 결과(RMSE = 0.027, MAPE = 0.007)는 거시경제 요인만을 고려한 모델 1-10(RMSE = 0.087, MAPE = 0.023) 보다 상대적으로 향상된 분석 결과를 보이고 있다. 이러한 결과는 엉클 블록 발생 예측에 거시 경제 요인 보다 직접적으로 관련이 있는 블록체인의 정보가 크게 기여한다는 것으로 볼 수 있다. 따라서, 거시경제 요인들을 고려했을 때 보다 블록체인의 정보만을 적용하여 엉클블록의 발생을 예측하는 것이 타당하고 할 수 있다.

마지막 세 번째 파트 분석에서는 거시 경제 요인과 블록체인 정보를 함께 적용하여 엉클블록의 발생을 예측하고자 하였다. 아래 <표 6>을 살펴보면 ANN은 SVM보다 모든 분석 모델에서 우월한 분석 결과를 제공하고 있다. 모델 3-1에서부터 모델 3-10까지의 분석은 거시경제 요인과 블록체인의 정보를 각각 위계적으로 추가하여 분석한 결과이며, 모델 3-8까지는 변수가 추가될 때마다 분석 성능이 향상되었으나 모델 3-9와 모델 3-10에서는 분석결과가 개선되지 않았다. 거시경제 요인과 블록체인 정보를 활용한 세 번째 파트 분석에서 모델 3-8은 우수한 분석 결과를 제공하고 있다 (RMSE = 0.030, MAPE = 0.008). 하지만 블록체인 정보만을 고려한 두 번째 파트 분석의 모델 2-9(RMSE = 0.027, MAPE = 0.007)보다는 상대적으로 부족한 분석 결과이다. 따라서, 엉클블록 발생을 보다 정확히 예측하기 위해서 거시 경제 요인과 블록체인 정보를 함께 고려하기 보다는 블록체인의 정보만을 고려하는 것이 타당하다고 할 수 있다.

〈표 5〉 블록체인 정보 활용 분석 결과

		모델 2-1	모델 2-2	모델 2-3	모델 2-4	모델 2-5	모델 2-6	모델 2-7	모델 2-8	모델 2-9
블록체인 내 거래 금액		√	√	√	√	√	√	√	√	√
성사된 거래 수			√	√	√	√	√	√	√	√
생성된 블록 수				√	√	√	√	√	√	√
블록 생성 난이도					√	√	√	√	√	√
블록의 데이터 크기						√	√	√	√	√
생성된 코인 양							√	√	√	√
거래 수수료								√	√	√
이더리움 가격									√	√
블록체인 활동 계정 수										√
ANN	RMSE	0.236	0.118	0.096	0.094	0.097	0.033	0.036	0.027	**0.027**
	MAPE	0.071	0.031	0.025	0.024	0.025	0.008	0.008	0.008	**0.007**
SVM	RMSE	0.232	0.133	0.105	0.104	0.104	0.102	0.100	0.100	0.100
	MAPE	0.064	0.033	0.027	0.026	0.026	0.026	0.025	0.025	0.025

〈표 6〉 거시경제 요인 및 블록체인 정보 활용 분석 결과

		모델 3-1	모델 3-2	모델 3-3	모델 3-4	모델 3-5	모델 3-6	모델 3-7	모델 3-8	모델 3-9	모델 3-10
S&P 500		√	√	√	√	√	√	√	√	√	√
DOW 30			√	√	√	√	√	√	√	√	√
Eurostoxx 50				√	√	√	√	√	√	√	√
NASDAQ					√	√	√	√	√	√	√
Crude oil						√	√	√	√	√	√
SSE							√	√	√	√	√
Gold								√	√	√	√
VIX									√	√	√
Nikkei 225										√	√
FTSE100											√
블록체인 내 거래 금액		√	√	√	√	√	√	√	√	√	√
성사된 거래 수			√	√	√	√	√	√	√	√	√
생성된 블록 수				√	√	√	√	√	√	√	√
블록 생성 난이도					√	√	√	√	√	√	√
블록의 데이터 크기						√	√	√	√	√	√
생성된 코인 양							√	√	√	√	√
거래 수수료								√	√	√	√
이더리움 가격									√	√	√
블록체인 활동 계정 수										√	√
ANN	RMSE	0.120	0.094	0.083	0.082	0.078	0.035	0.035	0.030	0.033	0.033
	MAPE	0.032	0.025	0.021	0.020	0.018	0.008	0.009	0.008	0.009	0.008
SVM	RMSE	0.175	0.126	0.093	0.092	0.092	0.090	0.088	0.087	0.086	0.086
	MAPE	0.049	0.030	0.021	0.021	0.021	0.021	0.020	0.020	0.020	0.019

V. 결 론

5.1 학문적 기여사항

본 연구의 학문적 기여사항은 다음과 같다. 첫 번째, 본 연구는 머신러닝 기법을 활용하여 블록체인 정보와 거시 경제 요인이 엉클블록 발생과 관련 있다는 사실을 발견하였다. 기존 연구에서는 블록체인에 대한 적용 방안이나 코인 가격 예측 등의 연구가 주로 수행되었다. 반면에 블록체인 시스템의 안정성과 지속적인 발전을 위해 엉클블록에 대한 연구가 필요한 상황임에도 불구하고 이에 대한 연구는 거의 수행되지 않았다. 학문적 갈증의 해소를 위해 본 연구는 이더리움 블록체인을 대상으로 엉클블록 발생을 예측 분석하였다. 본 연구는 블록체인 정보와 거시 경제 요인이 엉클블록 발생에 관련된다는 사실을 제공함으로써 향후 블록체인과 엉클블록 관련 연구에서 필요한 변수와 이론적 기반을 제공하였다. 또한, 블록체인의 안정적인 시스템 구현을 위한 엉클블록 발생의 탐색적 연구를 수행함으로써 정보시스템 분야에서 블록체인 연구에 기여할 수 있는 또 다른 방안을 제시하였다는 학문적 의미가 있다.

두 번째, 본 연구는 보다 정확한 엉클블록 발생 예측을 위하여 블록체인의 정보만을 활용하는 것이 타당하다는 사실을 제공한다. 본 연구의 발견은 거시 경제 요인과 블록체인의 정보가 엉클블록 발생에 관련되어 있다는 사실을 제공한다. 추

가적으로, 본 연구는 엉클블록의 발생 예측에 관련되는 변수들을 구체적이고 명확하게 밝히고자 하였다. 세 개의 파트로 구성된 본 연구의 분석은 블록체인의 정보만을 고려하였을 때 보다 적은 변수로 가장 좋은 분석결과를 얻을 수 있다는 사실을 제공한다(모델 2-9 예측 결과: RMSE = 0.027, MAPE = 0.007). 특히, 성사된 거래 수, 생성된 코인, 이더리움 가격은 엉클블록 발생 예측에 크게 기여한다는 사실을 발견하였다. 본 연구는 머신 러닝기법을 활용한 엉클블록 발생 예측에 적절한 변수들을 발굴해내고 변수 선택의 중요성을 제시하였다는 학문적 의미를 지닌다.

세 번째, 본 연구는 인공 신경망 머신 러닝기법이 엉클블록 발생 예측에 가장 우수한 성능을 제공한다는 사실을 발견하였다. 머신 러닝기법은 도메인에 따라 최적의 성능을 제공하는 머신 러닝기법이 다르게 나타날 수 있다(Wang *et al.*, 2014). 하지만 아직까지 블록체인의 엉클블록 발생 문제를 가장 잘 예측하는 머신 러닝기법은 확인되지 않았다. 본 연구는 선행 연구에서 대표적으로 활용되어온 인공 신경망과 서포트 벡터 머신 기법을 활용하여 엉클블록 발생을 예측하고자 하였다. 분석결과, 인공 신경망이 엉클블록 발생 예측에 가장 우수한 분석 결과를 제공한다는 사실을 발견하였다. 향후 연구에서는 엉클블록 관련 문제 분석에 있어 인공 신경망 머신 러닝기법을 적극적으로 활용해야 할 것이다.

5.2 실무적 기여사항

본 연구의 실무적 기여사항은 다음과 같다. 첫 번째, 블록체인 관련 기업은 블록체인 시스템 정보를 활용하여 엉클블록의 발생을 예측하고 사전에 엉클블록의 과도한 발생으로 인한 문제를 대비할 수 있을 것이다. 특히, 하드포크 현상의 조짐을 미리 예측하여 그에 따른 대응방안을 수립할 수 있을 것으로 기대한다. 블록체인 관련 기업에서는 엉클블록에 관련된 문제를 대비하기 위해 블록체인에서 성사된 거래 수, 생성된 코인, 코인 가격을 중요하게 모니터링 할 필요가 있어 보인다. 엉클블록의 과대 발생을 대비해 관련 있는 블록체인 요인들을 사전에 조정 및 관리하는 방안도 엉클블록으로 인한 문제를 줄이는 해결책이 될 수 있을 것이다.

두 번째, 블록체인 관련 기업들은 머신 러닝기법을 활용하여 엉클블록을 예측하고자 할 때 블록체인 정보와 인공신경망 알고리즘을 적극적으로 활용하여 높은 예측력을 얻을 수 있을 것이다. 거시경제 요인들을 고려하지 않고 블록체인정보만을 활용한 인공신경망은 기업에게 보다 신속하고 정확한 예측 결과를 제공할 수 있을 것이다. 이를 통해 불필요한 컴퓨터 연산이나 시간 손실을 줄일 수 있을 것으로 예상한다.

5.3 한계점 및 향후 연구방향

본 연구의 한계점 및 향후 연구 방향은 다음과 같다. 첫 번째, 본 연구는 이더리움 블록체인의 엉클블록만을 대상으로 연구를 진행하였다는 한계점이 있다. 블록체인은 비트코인, 대시, 라이트코인 등 다양한 블록체인 시스템이 존재한다. 이더리움이 2세대 블록체인으로 대표되고 있지만 다른 블록체인 시스템은 블록의 생성 주기, 코인의 발행량 등의 차이가 존재한다. 따라서 다양한 블록체인 시스템을 대상으로 엉클블록의 발생을 연구할 필요가 있다. 향후 연구에서는 이러한 부분을 참고하여 다양한 블록체인 시스템의 엉클블록 발생을 분석하고 시스템 간의 비교 분석을 진행하는 연구가 필요할 것이다.

두 번째, 본 연구는 인공 신경망과 서포터 벡터 머신만을 분석기법으로 활용하였다는 한계점을 가진다. 비록 선행 연구에서 인공 신경망과 서포트 벡터 머신 기법이 우수한 분석 결과를 제공하는 것으로 나타났으나 인공 신경망과 서포트 벡터 머신 이외에도 엉클블록 발생을 예측 분석할 수 있는 알고리즘은 충분히 존재한다. 하지만 본 연

구에서는 선행 연구에서 주로 사용되어온 인공신경망과 서포트 벡터 머신 만을 고려하였다. 향후 연구에서는 분석 결과의 성능을 높일 수 있는 앙상블 기법이나 여러 다른 분석 기법이 적용되어야 할 것이다. 이를 통해 엉클블록 발생 예측에 최적의 분석기법을 발견할 수 있을 것으로 기대한다.

참 고 문 헌

[1] 김진화, 남기찬, 이상종, "Support Vector Machine 기법을 이용한 고객의 구매의도 예측", *Information Systems Review*, 제10권, 제2호, 2008, pp. 137-158.

[2] Abraham, J., Higdon, D., J. Nelson, and J. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis", *SMU Data Science Review*, Vol.1, No.3, 2018, pp. 1-23.

[3] Antonopoulos, A. M., *Mastering Bitcoin: Unlocking Digital Cryptocurrencies*, O'Reilly Media Inc, 2014.

[4] Antonopoulos, A. M. and G. Wood, *Mastering Ethereum: Building Smart Contracts and Dapps*, O'Reilly Media, 2018.

[5] Breitman, K., "Op Ed: Why ethereum's hard fork will cause problems in the coming year", 2017, Available at https://bitcoinmagazine.com/articles/op-ed-why-ethereums-hard-fork-will-cause-problems-coming-year/.

[6] Ciaian, P., M. Rajcaniova, and D. A. Kancs, "The economics of Bitcoin price formation. Applied Economics", *Applied Economics*, Vol.48, No.19, 2016, pp. 1799-1815.

[7] Cortes, C. and V. Vapnik, "Support-vector networks", *Machine learning*, Vol.20, No.3, 1995, pp. 273-297.

[8] Jang, H. and J. Lee, "An empirical study on modeling and prediction of Bitcoin prices with bayesian neural networks based on blockchain information", *IEEE Access*, Vol.6, 2017, pp. 5427-5437.

[9] Kara, Y., M. A. Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange", *Expert systems with Applications*, Vol.38, No.5, 2011, pp. 5311-5319.

[10] Kim, Y. B., J. G. Kim, W. Kim, J. H. Im, T. H. Kim, S. J. Kang, and C. H. Kim, "Predicting fluctuations in cryptocurrency transactions based on user comments and replies", *PloS One*, Vol.11, No.8, 2016, e0161197.

[11] Kaytez, F., M. C. Taplamacioglu, E. Cam, and F. Hardalac, "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines", *International Journal of Electrical Power & Energy Systems*, Vol.67, 2015, pp. 431-438.

[12] Kristoufek, L., "BitCoin meets Google trends and wikipedia: Quantifying the relationship between phenomena of the Internet era", *Scientific reports*, Vol.3, 2013, p. 3415.

[13] Kuo, T. T., H. E. Kim, and L. Ohno-Machado, "Blockchain distributed ledger technologies for biomedical and health care applications", *Journal of the American Medical Informatics Association*, Vol.24, No.6, 2017, pp. 1211-1220.

[14] Li, Z., J. Kang, R. Yu, D. Ye, Q. Deng, and Y. Zhang, "Consortium Blockchain for secure energy trading in industrial internet of things", *IEEE Transactions on Industrial Informatics*, Vol.14, No.8, 2018, pp. 3690-3700.

[15] Mallqui, D. C. and R. A. Fernandes, "Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques", *Applied Soft Computing*, Vol.75, 2019, pp. 596-606.

[16] Mensi, W., K. H. Al-Yahyaee, and S. H. Kang, "Structural breaks and double long memory of cryptocurrency prices: A comparative analysis from Bitcoin and Ethereum", *Finance Research Letters*, Vol.29, 2019, pp. 222-230.

[17] Murphy, K. P., *Machine Learning: A Probabilistic Perspective*, Cambridge, MA, USA: MIT Press, 2012.

[18] Park, J. and T. Hong, "The prediction of DEA based efficiency rating for venture business using multi-class SVM", *Asia Pacific Journal of Information Systems*, Vol.19, No.2, 2009, pp. 139-155.

[19] Pati, J., B. Kumar, D. Manjhi, and K. K. Shukla, "A comparison among ARIMA, BP-NN, and MOGA-NN for software clone evolution prediction", *IEEE Access*, Vol.5, 2017, pp. 11841-11851.

[20] Wang, G., J. Sun, J. Ma, K. Xu, and J. Gu, "Sentiment classification: The contribution of ensemble learning" *Decision Support Systems*, Vol.57, 2014, pp. 77-93.

[21] Wang, P., "Pricing currency options with support vector regression and stochastic volatility model with jumps," *Expert Systems with Applications*, Vol.38, No.1, 2011, pp. 1-7.

[22] Wu, H., Li, Z., King, B., Ben Miled, Z., Wassick, J., and Tazelaar, J., "A distributed ledger for supply chain physical distribution visibility", *Information*, Vol.8, No.4, 2017, p. 137.

[23] Yao, J. and C. L. Tan, "Time dependent directional profit model for financial time series forecasting", *Neural Computing: New Challenges and Perspectives for the New Millennium*, Vol.5, 2000, pp. 291-296.

[24] Yue, X., H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare data gateways: found healthcare intelligence on Blockchain with novel privacy risk control", *Journal of Medical Systems*, Vol.40, No.10, 2016, p. 218.

[25] Zheng, Z., S. Xie, H. N. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: a survey", *International Journal of Web and Grid Services*, Vol.14, No.4, 2018, pp. 352-375.

Information Systems Review
Volume 22 Number 1
February 2020

A Study on Uncle Block Analysis of Blockchain Using Machine Learning Techniques

Han-Min Kim*

Abstract

Blockchain is emerging as a technology that can build trust between users participating in the system. As interest of Blockchain has increased, previous studies have mainly focused on cryptocurrency and application methods related to Blockchain technology. On the other hand, the studies on the stable implementation of Blockchain were rarely conducted. Typically, uncle block in the Blockchain plays an important role in the stable implementation of the Blockhain system, but no study was conducted on this. Drawing on this recognition, this study attempts to predict the uncle block of Blockchain using machine learning method, Blockchain information, and macro-economic factors. The results of artificial neural network and support vector machine analysis, Blockchain information and macro-economic factors contributed to the prediction of uncle block of Blockchain. In addition, artificial neural network using only Blockchain information provided the best performance for predicting the occurrence of uncle block. This study suggests ways to lead and contribute to Blockchain research in information systems filed.

Keywords: Blockchain, Uncle Block, Machine Learning, Prediction, Blockchain Information

* Ph.D Candidate, Business School, Sungkyunkwan University

◐ 저 자 소 개 ◑

김 한 민 (hanmin8809@gmail.com)

현재 성균관대학교 경영대학에서 경영 정보 시스템을 전공으로 박사과정 학생이다. 관심 연구 분야는 블록체인, 빅데이터 분석, 악성댓글, 온라인 일탈행동이다. Information Technology and People(IT&P), Asia Pacific Journal of Information Systems(APJIS), Journal of Information Technology Applications & Management (JITAM), 대한경영학회지, 디지털융복합연구 등의 저널에 논문을 게재하였다. 학술대회 활동으로 Americas Conference on Information Systems(AMCIS), 한국경영정보학회 학술대회 등에서 블록체인과 악성댓글을 주제로 학술 논문을 발표하였다.

논문접수일 : 2019년 05월 13일 **게재확정일 : 2019년 09월 24일**
1차 수정일 : 2019년 08월 26일