머신 러닝 기반의 비트코인 주소 분류 기법

이채현 O*, 고경찬*, 우종수**, 홍원기*

*포항공과대학교 컴퓨터공학과

**포항공과대학교 정보통신연구소

{chlee0211, kkc90, woojs, jwkhong}@postech.ac.kr

A Method of Machine Learning based Bitcoin Address Classification

Chaehyeon Lee O*, Kyunchan Ko*, Jongsoo Woo**, James Won-Ki Hong*

*Department of Computer Science and Engineering, POSTECH

**Graduation School of Information Technology, POSTECH

요 약

비트코인을 거래하기 위해서는 비트코인 주소가 필요한데, 이 주소는 실 사용자의 정보를 연결하지 않는다는 익명성을 띠고 있다. 이러한 익명성을 악용하여, 비트코인 네트워크에서 다양한 불법 거래들이 활발하게 일어나고 있고 피해가 막심하다. 이에 본 논문에서는 거래와 관련된 비트코인 주소의 특성을 파악하고 주소의 분류를 예측하는 방법론을 제안한다. 여러 서비스 (거래소, 마이닝 풀, 믹서, 겜블링, 암거래 시장 - 실크로드)에 활용된 트랜잭션을 카테고리 별로 수집하고, 수집된 트랜잭션으로부터 연관된 비트코인 주소와 80 개의 특징을 추출한다. 그리고 머신 러닝 모델을 이용해 특정 비트코인 주소가 어떤 카테고리에 속하는지 분류해보았고 최고 약 84%의 분류 정확도를 가짐을 확인했다.

I. 서 론

비트코인 [1]은 P2P 네트워크 구조의 탈 중앙화 시스템으로, 제 3 자의 개입 없이 암호 화폐의 거래가 가능하다. 참여자가 동일한 데이터를 유지함으로써 투명한 거래가 가능하며 데이터의 위/변조가 불가능하다는 특징이 있어 크게 주목받고 있다. 비트코인을 거래하기 위해서는 비트코인을 전송 받을 비트코인 주소가 필요하며, 한 명의 유저 또는 entity 는 여러 개의 주소를 소유할 수 있다.

그러나 이러한 비트코인 주소는 실 소유주의 현실 정보와 연결되지 않는 익명성을 띠고 있다. 이러한 익명성을 악용하여, 비트코인 네트워크에서 다양한 불법 거래들이 활발하게 일어나고 있고 피해가 막심하다 [2, 3]. 대표적인 암거래 사이트인 실크로드를 통해 불법 무기, 불법 마약, 도난품, 악성 코드 등이 비트코인을 이용해 거래되었으며 그 거래 금액은 사이트가 폐쇄되기 전까지 약 707M 달러에 달한다 [4]. 실제로 비트코인 네트워크 상에서는 다크넷을 통한 불법 물품 거래 뿐 아니라, 자금 세탁, 스캐밍과 같은 불법적인 활동들이 꾸준히 일어나고 있으며, 이러한 활동들은 암호 화폐 관련 법 제정을 저해하는 요소로 작용해왔다. 따라서 네트워크 상에서 일어나는 불법 거래들을 사전에 탐지할 수 있는 시스템이 필요하다. 비트코인, 이더리움 [5], 모네로 [6] 등은 다크넷 상에서의 불법 거래에 많이 사용되어 왔지만, 본 연구에서는 다크넷에서 가장 많이 사용된 비트코인에 초점을 맞춘다. 악성 유저들은 반복해서 불법 거래를 발생시키며, 그 과정에서 한 명의 유저는 여러 개의 비트코인 주소를 활용한다. 따라서 본 연구에서는 불법 거래를 탐지하기에 앞서, 특정 비트코인 주소들이 어떠한 목적으로, 어떠한 서비스를 위해 사용되었는지를 분류해보았다. 비트코인 거래와 관련한 비트코인 주소와 특징들을 추출하고, 지도 학습 기반의 머신 러닝 분류 모델 [7]을 이용해 특징들을 학습시킴으로써 비트코인 주소가 어느 클래스에 속하는지 파악해보았다.

II. 관련 연구

Toyota [8]는 트랜잭션 패턴을 분석하여 HYIP (High Yielding Investment Program)인지 아닌지를 식별하는 연구를 진행하였다. 수동적으로 HYIP 와 non-HYIP 주소를 식별하고 비트코인 주소와 연관된 트랜잭션의 수, 채굴된 블록의 수 등의 특징을 추출하였다. 비트코인 주소를 HYIP 또는 non-HYIP 로 라벨링 하여 supervised 학습을 통해 사이버범죄 집단을 분류하였다. 83%의 분류 정확도를 보여주었다.

Kanemura [9]는 다크넷 [10]과 관련된 비트코인 트랜잭션과 주소들을 분석하였고 하나의 Entity 에 속하는 여러 개의 주소들의 라벨을 결정하기 위해 voting-based 시스템을 제안했다. 다크넷 마켓과 관련한 73 개의 특징들을 추출하고 supervised 분류기를 이용해 데이터를 학습시킨 후, 특정 주소가 DNM 에 속하는 주소인지 non-DNM 에 속하는 주소인지 분류를 결정하였다. 논문에서 제안한 Majority voting 기반의 voting method 를 통해 81%에 해당하는 분류 정확도를 도출했다.

본 논문의 저자는 이전 연구를 통해, 비트코인 트랜잭션의 특징만을 이용해 불법 거래를 탐지하는 연구를 진행했다 [11]. 불법 활동과 연관된 비트코인 주소나 비

트코인 클러스터를 분류하는 연구는 사전에 여러 차례 진행되어 왔지만 트랜잭션의 특징만을 이용해 불법 거래를 탐지한 사전 연구를 찾을 수 없었기 때문에 트랜잭션으로부터 9 개의 특징을 추출하고, 하나의 라벨을 붙여 10 개의 특징을 이용해 머신러닝 분류 모델을 학습시켰다. 실험 결과 0.9 에 달하는 F1 score 를 얻을 수 있었지만 테스트 셋이 overfitting 되었을 가능성이 있고, 트랜잭션의 불법 여부를 결정짓기에 특징의 수가 너무 적다고 판단하였다.

따라서 이전 연구에 이어, 비트코인 트랜잭션의 특징에서 비트코인 주소의 특징으로 범위를 확장하였고, 특징의 수를 늘려 어떤 특징들이 분류 모델에 영향을 끼치는지 확인해보았다.

III. 비트코인 주소 분류 방법론

비트코인 주소의 카테고리를 분류하기 활용한 4 단계의 방법론을 소개한다. 1. 트랜잭션 수집, 2. 비트코인 주소 및 특징 추출, 3. 기계학습 분류 모델 학습, 4. 실험 및 검증 단계가 이에 해당하며 그림 1 과 같은 절차를 따른다.

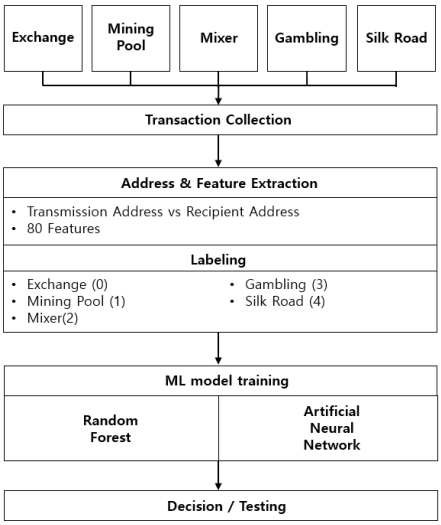

그림 1 불법거래 탐지 절차

1. 트랜잭션 수집

기계학습 모델을 구현하기에 앞서, 공개되어있는 포럼 사이트로부터 트랜잭션의 해시 리스트를 수집한다. WalletExplorer.com [12]은 트랜잭션의 해시를 카테고리 별(거래소, 마이닝 풀, 서비스 등)로 공개하고 있다. 본 논문에서는 믹서, 거래소, 겜블링, 마이닝 풀, 실크로드의 총 5 개 카테고리로부터 데이터를 수집하였다. 각 카테고리 별로 2016 년 1 월 1 일부터 2019 년 10 월 16 일까지의 데이터를 수집하였고, 실크로드의 경우는 사이트가 폐쇄되기 전까지의 데이터만을 포함한다.

2. 비트코인 주소 및 특징 추출

수집된 트랜잭션 해시로부터 하나 이상의 전송 주소와 하나 이상의 수신 주소를 추출할 수 있다. 수집한 트랜잭션 해시를 인자로 하여 JSON-RPC [13]를 요청한 결과 트랜잭션의 디테일한 정보를 비트코인 클라이언트 [14]로부터 얻어올 수 있다. 특정 비트코인 주소는 오직 전송 주소로만 사용되거나 수신 주소로만 사용되며, 전송과 수신을 위해 모두 사용되는 경우도 있다. 아래의 표 1 과 그림 2 는 카테고리 별 수집된 트랜잭션의 수와 추출한 수신/송신/총 비트코인 주소의 분포를 보여준다. 마이닝 풀과 믹서의 경우 상대적으로 적은 비율의 주소들이 추출되었고, 이를 통해 특정 주소가 해당 서비스를 위해 트랜잭션에 여러 번 등장했음을 유추해 볼 수 있다.

비트코인 주소를 추출한 후, 불법 거래들이 띠는 공통적인 패턴을 분석하기 위해 주소들로부터 특징을 추출한다. 28 개의 특징을 선정하였고, 일부 특징들은 평균값, 총합, 최소값, 최대값의 4 개의 값들을 포함한다. 특정 비트코인 주소는 우리가 수집한 트랜잭션에 여러 번 등장할 수 있으므로 동일한 비트코인 주소가 발견될 시 해당 값들을 업데이트한다. 전송 주소와 수신 주소 별로 특징이 추출되며, 0 의 값이 중복되는 것을 방지하기 위해 전송 주소의 경우엔 비트코인 수신과 관련한 필드들이 −1 로, 수신 주소의 경우엔 전송과 관련한 필드들이 −1 로 세팅된다. 다시 말해, 추출된 특징의 모든 값들이 −1 을 포함하지 않는다면 해당 주소는 전송과 수신 모두를 위해 사용된 주소이다. 4 개의 값(평균값, 총합, 최소값, 최대값)을 갖는 일부 특징들을 포함해 총 80 개의 특징을 추출하였고 각 카테고리 별로 0 에서 4 까지의 라벨을 부여하였다. 표 2 와 표 3 은 카테고리 별 라벨값과 추출한 특징에 대한 설명을 명시한다.

표 1 카테고리 별 추출된 거래 및 주소의 분포

Category	Transmission addresses	Recipient addresses	Total addresses	Transactions
Exchange	1,395,325	6,736,265	8,665,943	761,494
Mining Pool	218,476	1,036,143	1,375,327	325,800
Mixer	178,721	480,754	718,915	93,200
Gambling	726,210	3,960,029	5,345,783	752,300
Darknet (Silk Road)	704,376	938,730	2,305,872	956,186

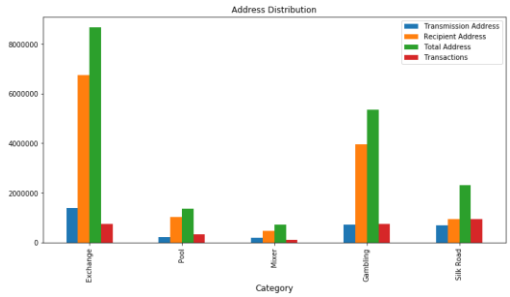

그림 2 카테고리 별 데이터 셋의 분포 비교

표 2 카테고리 별 라벨

Category	Label
Exchange	0
Mining Pool	1
Mixer	2
Gambling	3
Darknet (Silk Road)	4

3. 기계학습 분류모델 학습

수집된 주소들의 카테고리 분류를 위해 Random forest [15]와 DNN [16]두 가지의 기계학습 모델을 사용한다. 두 모델은 sklearn [17]과 tensorflow [18]를 이용해 구현하였으며 DNN 모델은 80 개의 특징을 학습시키기 위해 50 개의 노드를 가진 하나의 히든 레이어를 추가하였다.

표 3 추출된 특징과 Description

Name	Description	Etc
Bitcoin amount (transmit/receive)	Transmitted/Received bitcoin amount	
Total bitcoin amount (transmit/receive)	The amount of total bitcoin transmitted/received by the transaction associated with the transmission address	avg
Transaction fee (transmit/receive)	Transaction fees associated with bitcoin transmission/reception	sum
Sibling inputs/outputs (transmit/receive)	The number of sibling inputs/outputs	min
Sibling inputs/outputs_out/in (receive/transmit)	The number of outputs/inputs associated with bitcoin transmission/reception	max
Unique address (transmit/receive)	The number of unique transmission/receiving addresses	
Unique address_out/in (receive/transmit)	The number of unique receiving/transmission addresses associated with bitcoin transmission/reception	
Transaction Size (transmit/receive)	Transaction size associated with bitcoin transmission/reception	
Block Interval (transmit/receive)	The interval of the blocks related to the transmission/reception transaction	
Relevant transaction number (transmit/receive)	The number of transactions associated with the transmission/receiving address	
Lifetime (transmit/receive)	Life time of the transmission/receiving address	
First block (transmit/receive)	Block height where the transmission/receiving address first appeared	
Total transaction number	Total number of transactions associated with the address	
Total life time	Lifetime of the address	
Label	Classification of the address	

4. 모델 학습 및 테스트

학습이 완료되고 나면, test set 을 이용해 기계학습 모델이 제대로 구현되었는지 검증한다(그림 3). 구현이 잘 되었다면, 해당 모델을 활용해 주어진 특정 비트코인 주소가 어느 카테고리를 위해 사용되었는지 분류할 수 있다. 각 주소의 실제 라벨값과 예측값을 비교하여 모델이 얼마나 높은 정확도를 가지는지 확인할 수 있다.

lifetime_recv	lifetime_total	init_trns_block	init_recv_block	curr_trns_block	curr_recv_block	label	predicted
28610	28610	-1	577060	-1	577060	0	0
81636	61	520045	519984	520045	519984	2	2
-1	349145	256237	-1	256237	-1	4	4
371568	76	233367	233291	233367	233291	4	4
26277	26277	-1	466699	-1	492976	2	2
153551	7747	456638	448891	456638	448891	3	3
143460	143460	-1	458148	-1	458148	2	2
-1	380784	224025	-1	224025	-1	4	4
-1	46194	558509	-1	558509	-1	0	3
169877	169877	-1	432065	-1	432065	1	3
70557	70557	-1	531063	-1	531063	2	2
3716	3716	-1	548163	-1	551879	2	0
157921	157921	-1	444491	-1	444491	3	3
-1	33363	572057	-1	572057	-1	0	2
10887	10887	-1	595312	-1	595312	0	0
-1	378697	226127	-1	226127	-1	4	4
21590	21590	-1	584357	-1	584357	0	0
4763	4763	-1	438068	-1	442831	3	3
114093	114093	-1	493579	-1	493579	0	0
192875	192875	-1	408968	-1	408968	1	1

그림 3 테스트 셋의 예측 결과

IV. 실험 및 결과

그림 4 는 카테고리 별 추출된 트랜잭션과 주소의 분포를 나타낸다. 추출한 총 주소의 수는 18M 에 달하기 때문에 하드웨어 자원의 한계로 모든 데이터를 학습해 활용할 수 없다. 또한 각 카테고리 별로 불균형한 데이터 분포를 이루므로 실험을 위해서 데이터 셋을 카테고리 별로 랜덤하게 선정한다. 데이터 셋의 크기를 달리하여 여러 번 실험을 진행하며, 총 데이터 셋을 60:40 의 비율로 학습 데이터와 테스트 데이터로 나누었다.

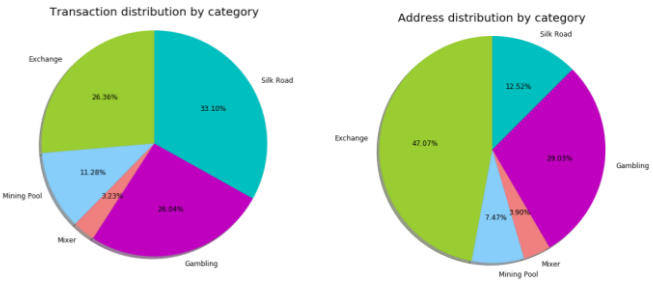

그림 4 카테고리 별 트랜잭션과 주소의 분포 비율

1. 특징 중요도

첫 번째 실험에서는 학습에 사용된 80 개의 특징들 중 분류 성능에 가장 영향을 많이 미치는 특징이 무엇인지 알아보았다. 각 카테고리 별로 2,000 개의 데이터를 랜덤하게 선택해 총 10,000 개의 데이터 셋을 이용하였고, 가장 중요도가 높은 Top 20 개의 특징을 조사하였다(그림 5). 실험 결과, 비트코인 수신과 관련한 특징들의 중요도가 높으며, 이는 데이터 셋에 수신 관련 주소가 많이 포함되었기 때문이라고 예측해볼 수 있다. 가장 영향을 많이 미친다고 판단되는 요소는 Lifetime 으로, 수집된 트랜잭션 셋에 걸쳐 특정 주소가 얼마나 오랫동안 해당 서비스를 위해 사용되었는지를 나타내는 지표이다. 특정 서비스가 동일한 주소를 반복해서 사용하는 경우 상대적으로 긴 Lifetime 을 가진다. 이 외에도 비트코인 주소가 수신한 총 비트코인의 양, 트랜잭션의 사이즈와 트랜잭션 수수료가 분류 정확도에 큰 영향을 미친다.

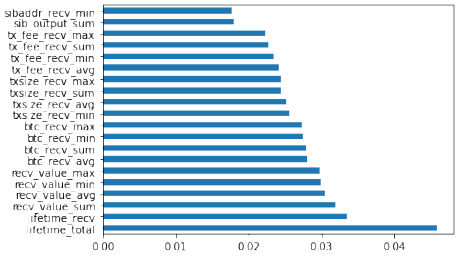

그림 5 카테고리 별 트랜잭션과 주소의 분포 비율

2. 분류 성능 비교

Random Forest Classifier 분류 성능 비교를 위해 각 카테고리 별로 1,000 개에서 200,00 개의 데이터를 랜덤하게 선택하여 여러 번의 실험을 반복하였다. 데이터 셋의 크기가 커질수록 분류 정확도가 높게 측정되었으며 최고 0.84 의 정확도를 보여주었다(표 4). 표 5 의 precision, recall, F1 score 는 카테고리 별로 분류가 얼마나 잘 되었는지 나타내는 지표이다. 실크로드와 관련된 주소의 경우, 모든 지표가 1.0 의 값을 가졌고 이는 실크로드와 관련된 주소들이 에러 없이 잘 분류되었음을 보여준다.

DNN 각 카테고리 별로 10,000 개에서 30,000 개의 데이터를 랜덤하게 선택하여 총 세 번의 실험을 반복하였다. 실험 결과 Random forest classifier 보다 상대적으

로 낮은 정확도와 F1 score 를 도출했다. 데이터 셋의 크기와 무관한 실험 결과를 보였으며 최고 정확도는 64% 이다(표 6). 믹서와 겜블링 관련 주소의 경우 단지 50% 정도만을 올바르게 분류하고 있음을 알 수 있었고, 그에 반해 실크로드 관련 주소는 Random forest classifier 와 같이 거의 정확히 분류해 줌을 확인해보았다(표 7).

표 4 Random forest classifier 의 정확도

Each data set	Accuracy
1,000	0.741
3,000	0.782
5,000	0.789
10,000	0.804
30,000	0.825
50,000	0.833
100,000	0.838
200,000	0.844

표 5 Random forest classifier 의 카테고리 별 정확도

Category	Precision	Recall	F1-Score
Exchange	0.82	0.74	0.78
Mining Pool	0.86	0.85	0.86
Mixer	0.78	0.86	0.82
Gambling	0.77	0.77	0.77
Silk Road	1.0	1.0	1.0

표 6 DNN 의 정확도

Each data set	Accuracy
10,000	0.646
20,000	0.620
30,000	0.614

표 7 DNN 의 카테고리 별 정확도

Category	Precision	Recall	F1-Score
Exchange	0.62	0.52	0.56
Mining Pool	0.77	0.56	0.65
Mixer	0.45	0.45	0.45
Gambling	0.51	0.46	0.48
Silk Road	0.99	0.98	0.99

V. 결론 및 향후 연구

본 연구에서는 머신 러닝 기반의 분류 모델을 이용해 여러 카테고리의 비트코인 주소들을 분류해보았다. 거래소, 마이닝 풀, 믹서, 겜블링, 암거래 시장(실크로드)에 활용된 트랜잭션을 카테고리 별로 수집하고, 수집된 트랜잭션으로부터 연관된 비트코인 주소와 80 개의 특징을 추출한다. 그리고 머신 러닝 모델을 이용해 특정 비트코인 주소가 어떤 카테고리에 속하는지 분류해보았고 최고 약 84%의 분류 정확도를 가짐을 확인했다. Random forest classifier 가 DNN 에 비해 상대적으로 높은 정확도를 보여주었고, 두 모델 모두 실크로드와 관련된 주소를 잘 분류해줌을 확인했다.

본 연구는 여러 개선사항들이 존재한다. DNN 성능 개선을 위해 다양한 머신 러닝 기법을 적용해 볼 예정이다. 히든 레이어의 수와 노드의 수를 조정하거나 80 개의 특징들 중 중요도가 높은 특징만을 이용해 재실험해 본다. 또한 다른 딥러닝 모델을 활용해 분류 정확도를 개선할 수 있다. 두 번째로, 특정 트랜잭션과 연관된 비트코인 주소들의 분류를 예측한 결과를 바탕으로 Majority voting 을 적용한다면, 더 나아가 해당 트랜잭션이 어떤 카테고리에 속하는지 판단할 수 있다. 실험 결과, 암거래 시장과 관련된 비트코인 주소가 거의 정확히 분류됨을 확인할 수 있었고, 이는 추후 트랜잭션의 불법 여부를 판단할 수 있는 가능성을 보여준다

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구 임 (No.2018-0-00539)

참 고 문 헌

[1] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008).
[2] Harvey, Campbell R. "Bitcoin myths and facts." (2014).
[3] Bitcoin Magazine. Bitcoin magazine: Bitcoin news, bitcoin charts, events. Available at https://bitcoi**nmagazine.com/articles/darknet-markets-cant-live-with-or-without-bitcoin.**
[4] Wikipedia. Silk road (marketplace). Available at https://en.wikipedia.o*rg*/wiki/Silk_Road_(**marketplace).**
[5] Gavin Wood et al. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow pape*r, 151(2014):1–32, 2014.
[6] monero. Zero to monero. Technical report, 2018. Available at https: **//www.getmonero.org/library/Zero-to-Monero**-1-0-0.pdf
[7] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learn ing: *A* review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160:3–24, 2007.
[8] K. Toyoda, T. Ohtsuki and P. T. Mathiopoulos, "Identification of High Yielding Investment Programs in Bitcoin via Transactions Pattern Analysis," *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Singapore, 2017, pp. 1-6.
[9] Kota Kanemura, Kentaroh Toyoda, and Tomoaki Ohtsuki. Identification of darknet markets' bitcoin addresses by voting per-address classification results. In *2019 IEEE International Conference on Blockchain and Cryp tocurrency (ICBC)*, pages 154–158. IEEE, 2019.
[10] Wikipedia. Darknet market. Available at https*://en*.m.wi**kipedia.org/wiki/Darknet_market.**
[11] LEE, Chaehyeon, et al. Toward Detecting Illegal Transactions on Bitcoin Using Machine-Learning Methods. In: *International Conference on Blockchain and Trustworthy Systems*. Springer, Singapore, 2019. p. 520-533.
[12] Walletexplorer: smart bitcoin block explorer. Available at **https:*//www*.walletexplorer**.com*/*.
[13] Bitcoin.org. Bitcoin core json apis. Available at https://bitcoin.org/e*n/***developer-reference#bitcoin-core-apis.**
[14] Bitcoin core. Available at https://bitcoin.o*rg/en/***bitcoin-core/.**
[15] Mahesh Pal. Random forest classifier for remote sensing classification. *In ternational Journal of Remote Sensing*, 26(1):217–222, 2005.
[16] Jacek M Zurada. *Introduction to artificial neural systems*, volume 8. West publishing company St. Paul, 1992.
[17] scikit-learn: Machine learning in python. Available at **https://scikit-learn.org/.**
[18] tensorflow: An end-to-end open source machine learning platform. Available at https://w**ww.tensorflow.org/?hl=en.**