

온라인 쇼핑몰 데이터를 활용한 판매동향 분석 시스템

차승연¹, 김강렬², 러비나스레스터³, 김영주⁴, 최종명^{5*}
¹²³목포대학교 컴퓨터공학과 학부생, ⁴⁵목포대학교 컴퓨터공학과 교수

Open Market Sales Trend Analysis System Using Online Shopping Mall Data

Seung-yeon Cha¹, Kang-ryeol Kim², Labina Shrestha³, Yeong-ju Kim⁴, Jongmyung Choi^{5*}

¹²³Student, Department of Computer Engineering, Mokpo National University

⁴⁵Professor, Department of Computer Engineering, Mokpo National University

요약 인터넷의 발전으로 온라인 쇼핑이 활성화 되면서 소비자들의 구매 형태가 기존의 대면 구매방식에서 온라인 구매방식으로 변하고 있다. 이에 수많은 판매자가 쇼핑몰로 유입되었고, 판매자들 간의 경쟁이 매우 치열한 실정이다. 따라서 쇼핑몰 내 판매자들은 소비자의 구매 패턴 및 제품 판매동향을 분석하여 합리적인 마케팅 전략을 세울 필요가 있다. 본 논문에서는 오픈 쇼핑몰에서 판매업자가 동일 제품을 판매하는 경쟁사의 제품 가격, 평점, 판매수량을 시간대 별로 분석하여 소비자의 구매 패턴을 파악했다. 또한 수집된 정보들을 차트로 가시화하여 자사와 경쟁사의 판매동향을 쉽게 비교할 수 있도록 하였다. 위 시스템을 사용하면 분석된 구매패턴을 통해 판매수량을 예측할 수 있고 판매동향을 파악하여 상품의 합리적인 가격 선정이 가능하다.

주제어 : 빅 데이터, 온라인 쇼핑, 마케팅 전략, 판매 동향

Abstract As online shopping is activated by the development of the Internet, consumers' purchase form is changing from the traditional face-to-face purchase method to online purchase method. Many sellers have flowed into shopping malls, and competition among sellers is very intense. Therefore, sellers in shopping malls need to establish rational marketing strategies by analyzing consumer purchase patterns and product sales trends. In this paper, we analyzed the purchase price of consumers by analyzing the product price, rating, and sales quantity of competitors who sell the same product in open shopping malls by time zone. In addition, the collected information was visualized in a chart so that the company's and competitors' sales trends could be easily compared. Using the above system, it is possible to predict the sales volume through the analyzed purchasing pattern and to select the reasonable price of the product by grasping the sales trend.

Key Words : Big Data, Online Shopping, Marketing Strategy, Sales Trend

1. 서론

농식품유통교육원 유통연구소에 따르면, 2017년 2월

국내 농산물 소매 거래액 385조 원 중 대형마트 53조 원, 온라인 쇼핑 52조 원, 슈퍼마켓 38조 원, 백화점 30조 원, 편의점 20조 원 순으로, 전자상거래의 보편화로

본 과제(결과물)는 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 사회맞춤형 산학협력 선도대학(LINC+) 육성사업의 연구 결과입니다.

*교신저자 : 최종명(jmchoi@mokpo.ac.kr)

접수일 2019년 9월 20일 수정일 2019년 11월 28일 심사완료일 2019년 12월 11일

인해 온라인 쇼핑이 매우 큰 규모를 차지하고 있다[1].

또한, 2019년 2월 국내 온라인 쇼핑 거래액은 9조 5,966억 원으로 전 년 동월대비 16.4%가 증가했고 이러한 현상은 온라인 쇼핑이 오프라인보다 다양한 제품 및 서비스의 정보를 신속하게 획득할 수 있으며, 동일한 제품일지라도 다른 온라인 상점이 제공하는 가격과의 비교가 신속하게 이루어지는 공간이기 때문이다[2].

최근에는 온라인 쇼핑물 플랫폼에 제삼의 판매자들이 참여하는 오픈 마켓 플랫폼의 비중이 커지고 있다. 하지만 오픈 마켓은 품목 별로 수많은 판매처가 존재하기 때문에 기존 업자들 간의 경쟁이 치열하고, 트렌드의 흐름을 따라 창업을 하려는 이에게는 진입장벽이 높다[3].

더불어, 실시간으로 변화하는 가격과 판매량으로 인해 효율적인 상품 전략을 세우기 어려운 문제점이 있다. 이를 해결하기 위하여 본 논문에서는 오픈 마켓 쇼핑물에서 제품정보들을 수집하고, 소비자 구매 패턴 및 판매동향을 분석하여 사용자가 보기 쉽게 가시화하여 보여주는 시스템을 제안한다. 이 시스템을 이용하면 분석된 구매패턴을 통해 판매수량 예측이 가능하고, 자사와 경쟁사의 판매 동향 분석을 통해 합리적인 가격과 소비자 중심의 상품구성을 통해 효과적인 마케팅 전략을 수립할 수 있다. 그리고 경쟁사의 판매 상황을 분석함으로써 경쟁력 확보가 가능하다. 이러한 결과로 제품의 가격변동이 최소화되면 농가의 효과적인 품질 및 재고 관리가 가능해지고 생산자의 소득까지 안정화될 것으로 기대된다.

논문은 2장에서 관련 연구를 소개하고, 3장에서는 제안하는 시스템에 대해서 설계 및 정보를 시각화하는 방법을 제시한다. 4장에서는 시스템의 실험 데이터와 5장에서는 결론을 밝힌다.

2. 관련 연구

제품의 가격변화와 판매상황을 분석하기 위한 다양한 연구들이 아래와 같이 선행되었다.

[4]의 논문에서는 경영관리에 있어 중요한 가격변화에 따른 판매량의 동적변화를 모형화하고 향후 가격변화에 따른 판매량 변화를 예측하였고 [5]의 논문에서는 국내 판매율 1, 2, 3위를 달리고 있는 3대 오픈마켓인 11번가와 G마켓, 그리고 옥션이 가지고 있는 고유 브랜드의 컬러가 제한된 스마트폰의 화면 속에서 어떻게 구현되고 있는지를 연구하였다.

또한, 오픈마켓의 데이터를 분석하기 위한 유사한 연

구들은 꾸준히 진행되고 있으며 대표적인 연구들로는 판매 빅데이터 분석[3], SNS 분석[6], 리뷰 분석[7] 등이 있다.

[3]의 논문에서는 오프라인 마켓의 POS 판매 데이터를 빅데이터 시스템으로 구현하여, 손익분석과 재고관리 등에 집중되었던 기존 판매 시스템을 보완하고, 고객의 소비 패턴 파악, 선호도 분석, 수요에 대한 사전 예측이 가능하도록 하는 시스템을 제시하였다.

[6]의 연구는 발작물의 판매 유형에 대한 분석을 위해서 뉴스, 블로그, 인터넷 카페 등의 SNS를 분석하여 발작물의 소비 선호도 및 패턴을 파악하는 연구를 진행하였다. 발작물의 품목 및 생산자단체에 대한 키워드를 검색하고, 이를 바탕으로 감성 분석과 연관어 분석을 실시하여 공급 측면에서의 전략적 대응을 마련할 수 있도록 하였다.

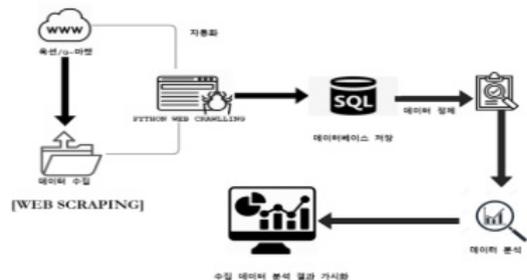
[7]의 연구에서는 사용자 리뷰를 바탕으로 모바일 소셜커머스와 오픈마켓의 사용자 이용경험을 비교 분석하는 연구를 수행하였다. 오픈마켓의 리뷰를 수집하고, 단어들에 유용성과 편리성 토픽으로 분류한 뒤 감정분석과 동시출현단어분석을 수행하고, 이를 바탕으로 오픈마켓의 사용상의 문제점 등을 개선할 수 있는 방법을 제시한다.

기존 연구들은 판매 정보 혹은 사용자의 감성에 관련된 분석이 주를 이뤘지만, 본 연구에서는 경쟁사의 가격을 실시간으로 분석하고, 이를 바탕으로 가격 정책을 세울 수 있다는 점이 기존 연구와 큰 차이점이다.

3. 제안 방법

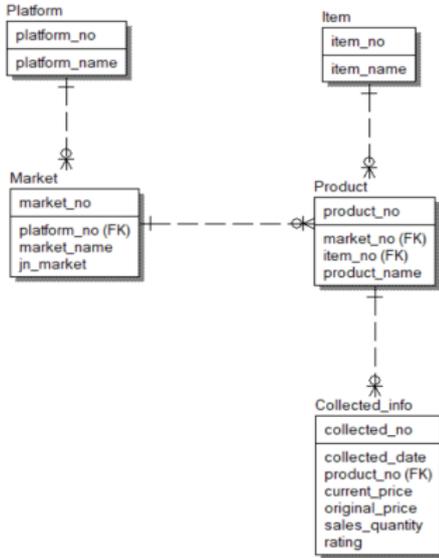
3.1 시스템 설계 구성

오픈마켓에서 경쟁사에 대해서 가격 경쟁력을 갖추기 위해서는 주기적으로 정보를 수집 및 저장하고, 이를 분석 및 가시화해서 경영진이 판단할 수 있도록 해야 한다. 따라서 시스템은 그림 1과 같은 형태로 데이터 수집 모듈, 데이터 분석 모듈, 데이터 가시화 모듈로 구성된다.



[Fig. 1] System Architecture

데이터 수집 모듈을 정기적으로 오픈마켓 플랫폼에서 정보를 수집하고, 이를 데이터베이스에 저장한다. 수집한 데이터는 그림 2와 같은 ER 다이어그램에 따라 각 테이블에 저장한다.



[Fig. 2] Logical ERD

데이터 분석 모듈에서는 저장된 데이터 간에 상관관계를 확인하여 상관관계가 없는 데이터들을 제거 한다. 그리고 데이터 필드 별로 수집된 시간대끼리 그룹화 하여 평균을 낸다. 이 과정을 통해 소비자의 구매 패턴을 알 수 있다.

데이터 가시화 모듈은 분석된 데이터들의 결과를 차트로 표현하여 사용자가 직관적으로 판매동향 및 구매패턴을 파악할 수 있도록 한다.

3.2 데이터 분석

본 연구는 매출에 가장 연관성 있는 판매수량을 기본으로 하여 분석한다. 데이터 필드 별로 수집된 시간대끼리 그룹화 하여 평균을 낸 뒤 시간대별 평균 판매수량, 평점 대비 판매수량, 할인가 대비 판매수량을 분석한다. 이 때 평점 대비 판매수량과 할인가 대비 판매수량은 상관관계를 알아내기 위해 상관분석, 회귀분석, 분산분석 [8]을 이용한다.

3.2.1 상관분석

데이터 간에 선형적 또는 비선형적 관계를 갖고 있는

지 분석하기 위해 상관 분석 한다. ρ 를 사용하여 상관관계의 정도를 예측한다.

3.2.2 회귀분석

데이터 사이의 인과관계를 규명하기 위해 회귀분석 한다. (할인가, 판매수량), (평점, 판매수량)의 함수관계를 밝히고 판매수량을 예측할 수 있다.

3.2.3 분산 분석

분산 분석을 통해 데이터들 비교하고 데이터 내의 분산, 총 평균과 각 데이터의 평균의 차이에 의해 생긴 데이터 간 분산의 비교를 통해 만들어진 F-분포를 이용하여 가설검정을 판단한다.

3.3 데이터 가시화

분석된 데이터를 가시화하기 위하여 Google Chart[9]를 사용한다. Google Chart는 다양한 차트 라이브러리 제공한다. 사용자에게 제공되는 차트 종류는 표 1과 같다.

<Table 1> Types of Charts Available for Users

Charts	Uses
Average Sales Quantity by time	Shows at which time of day the product is sold most.
Sales Quantity vs. Sale Price	Shows how the discount price affects the quantity sold.
Sales Quantity vs. Rating	Shows how rating affects sales quantity.
Change in Sales Volume	Shows the change in quantity of sales on a daily, weekly and monthly basis.
Change in Rating	Shows the change of ratings on daily, weekly or monthly basis.

4. 실험 및 결과

4.1 시스템 구현

데이터 수집 모듈에서는 온라인 쇼핑몰 A사의 J 판매업체의 제품의 상품 고유번호, 판매업체 이름, 상품명, 정가, 할인가, 판매수량, 평점을 1시간 마다 크롤링하여 데이터베이스에 저장한다. 만약 할인가가 적용되지 않았다면 정가를 할인가로 수집한다. 또한 A사 내에서 같은 품목을 판매하는 경쟁사의 판매동향도 분석하기 위해 경쟁회사의 데이터도 같은 항목으로 수집한다. 경쟁사 제품은 상품 검색 시 그 시간대에 가장 상단에 위치한 8개의

제품이다. 데이터 수집 모듈은 파이썬 웹 크롤링 기술을 사용한다. 필요한 라이브러리는 다음과 같다.

<Table 2> Libraries Used For Data Crawling[10]

Library	Uses
BeautifulSoup	Python's library used for parsing data easily from HTML and XML files.
urllib	Library used to get the desired web page before using HTML parser using BeautifulSoup library.
pymysql	Library used to connect python and MySQL.

4.2 실험 데이터 분석 결과

온라인 쇼핑몰 A 사의 J 판매업체의 제품 중 쌀, 토마토, 마늘의 데이터를 수집하였다. 또한 경쟁업체는 쌀, 토마토, 마늘 검색 시 그 시간대에 가장 상단에 위치한 8개의 제품으로 선정했다. 2019-11-11부터 2019-12-02까지 약 한 달 정도 수집하였으며 총 13,338건 데이터가 수집되었다.

4.2.1 J 판매업체 분석 결과

수집된 데이터를 필드 별로 수집 시간끼리 정렬한 후 부분 합하여 평균을 내었다. 다음은 분석 결과를 차트로 가시화한 결과이다.



[Fig. 3] Tomato Time vs Purchase Quantity

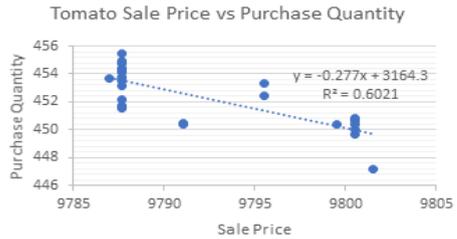


[Fig. 4] Garlic Time vs Purchase Quantity

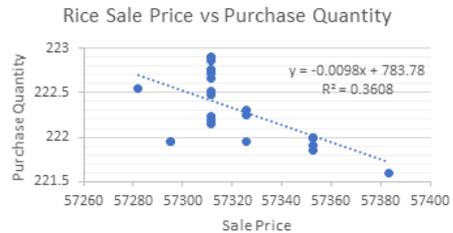


[Fig. 5] Rice Time vs Purchase Quantity

그림 3, 4, 5는 시간 별 평균 판매수량을 나타내는 차트이다. X축은 시간, Y축은 판매수량이다. 3개의 제품 모두 비슷한 양상을 보이며 16~20시에 가장 많이 판매된다.



[Fig. 6] Tomato Sale Price vs Purchase Quantity

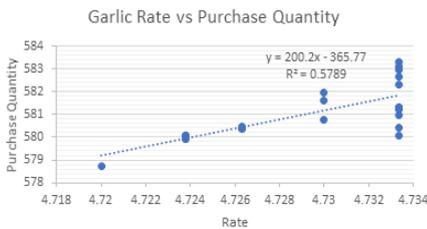


[Fig. 7] Rice Sale Price vs Purchase Quantity

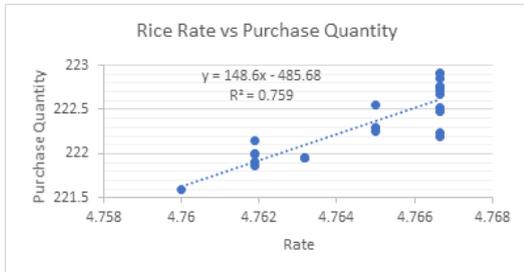
그림 6, 7은 할인가 대비 판매 수량을 분석한 그래프이다. 그래프를 통해 판매 수량과 할인가는 음의 상관관계를 가지고 있는 것을 볼 수 있다. 즉, 할인이 적용된 가격이 낮을수록 판매수량이 증가한다. 마늘은 가격이 변동하지 않아 분석되지 않았다.



[Fig. 8] Tomato Rate vs Purchase Quantity



[Fig. 9] Garlic Rate vs Purchase Quantity



[Fig. 10] Rice Rate vs Purchase Quantity

그림 8, 9, 10은 평점 대비 판매 수량을 분석한 그래프이다. 그래프를 통해 판매 수량과 평점은 양의 상관관계를 가지고 있는 것을 볼 수 있다. 즉, 제품의 평점이 높을수록 판매수량이 증가한다.

<Table 3> Correlation Analysis of Tomato

	Rating	Sale Price	Pur. Qty.
Rating	1		
Sale Price	-0.82869	1	
Pur. Qty.	0.887293	-0.77592	1

<Table 4> Correlation Analysis of Garlic

	Rating	Sale Price	Pur. Qty.
Rating	1		
Sale Price	#DIV/0!	1	
Pur. Qty.	0.76083389	#DIV/0!	1

<Table 5> Correlation Analysis of Rice

	Rating	Sale Price	Pur. Qty.
Rating	1		
Sale Price	-0.68791	1	
Pur. Qty.	0.871211	-0.60064	1

상관관계 분석 결과를 보면 평점과 판매수량은 강한 양의 관계, 할인가와 판매수량은 강한 음의 관계가 있다는 것을 알 수 있다.

<Table 6> Regression Results of Tomato and SalePrice

	Deg. of Freedom	Sum of Sq.	Mean Sq.	F Ratio	Sig. F
Reg.	1	60.2102	60.210	33.284	8.37E-06
Residual	22	39.7974	1.8089		
Tot.	23	100.007			

<Table 7> Regression Results of Rice and SalePrice

	Deg. of Freedom	Sum of Sq.	Mean Sq.	F Ratio	Sig. F
Reg.	1	1.2425	1.2425	12.416	0.00191
Residual	22	2.2016	0.1000		
Tot.	23	3.4441			

<Table 8> Regression Results of Tomato and Rating

	Deg. of Freedom	Sum of Sq.	Mean Sq.	F Ratio	Sig. F
Reg.	1	78.7350	78.7350	81.427	7.56E-09
Residual	22	21.2726	0.96694		
Tot.	23	100.007			

<Table 9> Regression Results of Garlic and Rating

	Deg. of Freedom	Sum of Sq.	Mean Sq.	F Ratio	Sig. F
Reg.	1	18.980	18.980	30.240	1.58762E-05
Residual	22	13.808	0.6276		
Tot.	23	32.788			

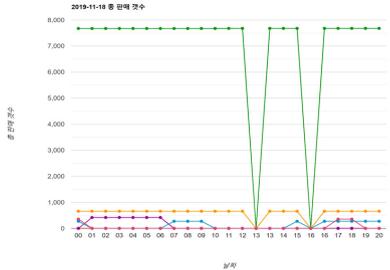
<Table 10> Regression Results of Rice and Rating

	Deg. of Freedom	Sum of Sq.	Mean Sq.	F Ratio	Sig. F
Reg.	1	2.6141	2.614	69.289	3.03472E-08
Residual	22	0.8300	0.037		
Tot.	23	3.4441			

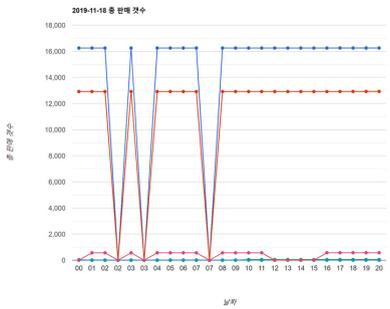
표 6~10의 F검정에서 p값이 거의 0에 가까운 값이므로 귀무가설을 기각할 수 있다.

4.2.2 경쟁업체 판매동향 분석 결과

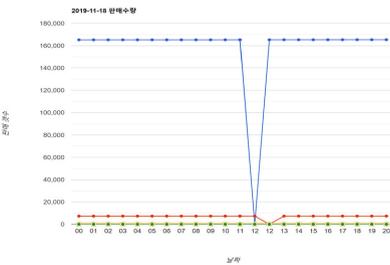
다음과 같이 판매수량과 평점의 변화 추이를 라인 차트로 표현하여 자사와 경쟁업체의 판매동향을 비교하여 분석할 수 있다.



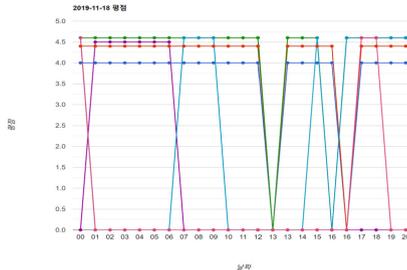
[Fig. 11] Purchase Changes in No.of Tomato



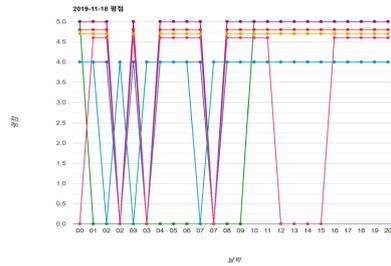
[Fig. 12] Purchase Changes in No.of Garlic



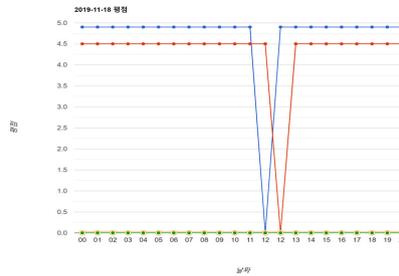
[Fig. 13] Purchase Changes in No.of Rice



[Fig. 14] Rating Changes in Tomato



[Fig. 15] Rating Changes in Garlic



[Fig. 16] Rating Changes in Rice

5. 결론

본 논문에서는 온라인 쇼핑몰 A사에서 공개된 판매수량, 가격, 평점 등을 수집하여, 웹 페이지 상에서 자사 및 경쟁사의 제품 판매동향 및 소비자 구매 패턴 분석을 가능케 하는 시스템을 제안하였다.

시스템의 결과를 토대로 마케팅 전략을 세우면 쉽게 경쟁사와 비교하며 제품의 합리적인 가격을 선정할 수 있고 소비자가 만족스러워 하는 상품 구성이 가능할 것이다. 또한 이 시스템이 지속적으로 사용되어 가격변동이 최소화 된다면, 농가의 효과적인 재고 관리 및 상품 품질 관리가 가능해지고 생산자의 소득도 변동 없이 안정적이게 될 것으로 기대된다.

REFERENCES

- [1] H.Y.Kim, "Current Status and Implications of Direct Commercial Trading in Private Social Commerce and Open Market", Agro-Fisheries & Food Trade Corp., Suwon, KR, Tech. Rep. In-depth-2017-000, 2017.
- [2] M.R.Kim and S.S.Kim, "Continuous Purchase Intention through Online Open Market Platform: Multidimensional Approach of Perceived Value in Purchasing Process

and Transaction Satisfaction", The Journal of Internet Electronic Commerce Research, Vol.19, No.3, pp.149-172, 2019.

- [3] H.Y.Lee, K.R.Park and D.H.Kim, "A Study on Possible Construction of Big Data Analysis System Applied to the Offline Market", The Journal of Digital Convergence, Vol.14, No.9, pp.317-323, 2016.
- [4] S.H.Park, J.D.Jun, "Estimation and Forecasting of Dynamic Effects of Price Increase on Sales Using Panel Data", The Journal of Korean Operations Research and Management Science Society, Vol.31, No.2, pp.157-167, 2006.
- [5] H.J.Lee and Y.L.Hong, "A Study on the Adaptation of Brand Colors on the Domestic Smartphone Open Market Apps Focusing on the 11st Street, G-Market, Auction", A Journal of Brand Design Association of Korea, Vol.13, No.2, pp.249-262, 2015.
- [6] K.S.Chae, B.Y.Kim and S.H.Min, "Analysis of Field Crop Consumers Using SNS Big Data", KREI, Naju, KR, Rec. Rep. R834-2, Dec. 2017.
- [7] S.H.Chae, J.I.Lim and J.Y.Kang, "A Comparative Analysis of Social Commerce and Open Market Using User Reviews in Korean Mobile Commerce", The Journal of Intelligence and Information Systems, Vol.21, No.4, pp.53-77, 2015.
- [8] Statistics, <https://en.wikipedia.org/wiki/Statistics>
- [9] Google Chart, <https://en.wikipedia.org/wiki/Statistics>
- [10] H.C.Ko, M.S.Kim, S.B.Lee and H.W.Lee, "Django based ChatBot System Using KakaoTalk API", KOITS, Vol.4, No.1, pp.31-36, 2018.

차 승 연(Seung-Yeon Cha) [준회원]



■ 2017년 3월 ~ 현재 : 목포대학교 컴퓨터공학과(학부 3학년)

<관심분야>
데이터베이스, 웹 프로그래밍

김 강 렬(Kang-Ryeol Kim) [준회원]



■ 2017년 3월 ~ 현재 : 목포대학교 컴퓨터공학과(학부 3학년)

<관심분야>
데이터베이스, 빅 데이터

러비나스레스터(Labina Shrestha) [준회원]



■ 2016년 3월 ~ 현재 : 목포대학교 컴퓨터공학과(학부 4학년)

<관심분야>
빅 데이터, 데이터베이스

김 영 주(Yeong-Ju Kim) [정회원]



- 2005년 2월 : 조선대학교 수학전산통계학부 이학사.
- 2008년 2월 : 목포대학교 정보컴퓨터교육 석사.
- 2017년 2월 : 목포대학교 컴퓨터공학 박사.
- 2017년 9월 ~ 현재 : 목포대학교 컴퓨터공학과 강사

<관심분야>
기계학습, 스마트양식, 풍력발전예측, 시계열분석

최 종 명(Jongmyung Choi) [중신회원]



- 2003년 8월 : 숭실대학교 컴퓨터학과 박사
- 2010년 8월 ~ 2011년 12월 : 미국 조지아공대 방문연구원
- 2004년 3월 ~ 현재 : 목포대학교 컴퓨터공학과 교수

<관심분야>
소셜네트워크, 헬스케어, HCI