International Journal of Information Technology, Research and Applications (IJITRA)
Vol.1, No.2, August 2022, pp. 9~15

Classification-based Collaborative filtering: A Machine Learning Recommendation System

Mahesh T R[1], V Vivek[1], Vinoth Kumar[1]
[1]Department of Computer Science and Engineering, JAIN (Deemed-to-be University), Bengaluru, India

Article Info

Article history:

Received June 04, 2022
Revised August 07, 2022
Accepted August 08, 2022

Keywords:

Recommendation Systems
Collaborative Filtering
Classification
Linear Regression

ABSTRACT

Social networking platforms like, Twitter, Face book etc., have now emerged as a major forum for the application of Recommended Systems (RSs). Of course these famous sites are taken as the main source of people-related knowledge and thus to be a great choice for exploiting modern and creative approaches to the recommendation, backing the old methods, in order to improve accuracy It was thought that helping users cope with the issue of data overload was the original role of information retrieval systems or search engines, but what separates recommended systems from the existing search engines is the requirements of personalized useful and interesting. The "intelligence" aspect is what suggests more interesting and useful. Intelligence is one of the main routes of personalization to know the interests of the user, anticipate the unknown favorites of the user, and eventually provide suggestions by matching the question and the content beyond a basic search. This article provides simple approaches to Recommendation Systems, provides recommendation for similar items based on the correlation and classification methods of machine learning to build a collaborative filtering system by making use of Logistic Regression model.

Corresponding Author:

Mahesh T R
Department of Computer Science and Engineering
JAIN (Deemed-to-be University)
Bengaluru
India
Email: trmahesh.1978@gmail.com

1. INTRODUCTION

The main objective of RS is to figure out as well as to recommend or suggest items that a user is very much interested in [1]. There are various examples of recommendation engines available like Amazon, Netflix, and Apple music. Recommendation systems are used for product recommendations such as Amazon and Etsy, in Netflix RS is used for movie recommendations, if we are music lovers then Apple music uses its recommendation system for music recommendations and there are social recommendation systems like face book, Instagram and LinkedIn [2-4].

Collaborative filtering systems recommend items based on crowdsourced information about users' preferences for items. There are two approaches in collaborative filtering approach [5-8]. They are a) user-based b) Item based collaborative filtering.

1.1 Item-based RS

The figure 1 shows Item based Collaborative filtering method. When we go to e-commerce web sites, we see items recommended to us are items that people who like these items also liked these items like a. b or c product. Item based approach generates recommendations based on the similarities between items with respect to user ratings about those items [9-13]. In figure 1, it is clearly shown that user B and user D bot have high ratings that is 4 stars to the cell phone and the cell phone case. The user A also likes the cell phone, so based on the similarities between user B and user D, let us recommend call phone case also to user A.

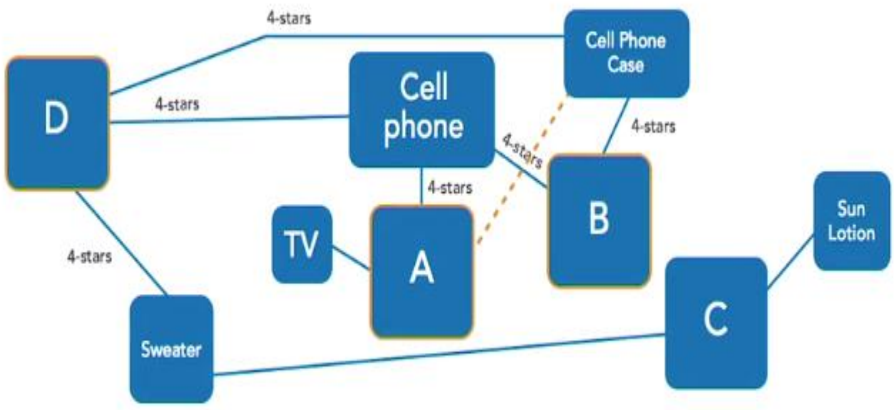

Figure 1: Item-based RS

1.2 User-based Recommendation Systems

The user-based Recommendation systems recommend items depending on similarities between the users [14-16]. For example, in e-commerce web site, recommended item to purchase, the recommendation will come in terms of customers who have similarities of a, b or c products. In table1, four different users along with their age, net worth of life insurance policy and marital status is given.

User	Age	Net worth ($)	Marital Status
User A	32	35,000	Divorced
User B	64	260,000	Married
User C	20	6,000	Single
User D	70	135,000	Married

Table 1: User-based recommendation System along with user ratings

If we observe in table 1 with respect to age, net worth and marital status, user B and D are quantitatively most similar. Based on known user attributes, we know that user B is like user D. User D really likes his life insurance policy, so let us recommend it to user B. This is an example for how user-based recommendation system may work.

Content-based Collaborative recommenders recommend the items based on the features as well as how similarities are to those features to other items in the data set [17-20]. Consider the data set shown in table2.

City Name	Average Temperature (F) in winter	Average cost of Living($)	Average Wi-Fi Speed(MBPS)
John, Texas	65.2	2102	175
Nick, Washington	50.7	3456	50

Table 2: Content-based Collaborative recommender system

From Table 2, we can observe that John and Abraham are more like one another compared to Nick of Washington. So, we can conclude that the user who loves Abraham also loves John based on the similarities between temperature, cost of living and Wi-Fi speed at both the places.

2. POPULARITY-BASED RECOMMENDER SUSTEMS

Popularity-based recommenders are very primitive type of Collaborative filtering where items are suggested based on how popular these items among others [21-24]. In table 3, place represents the item that we are recommending, and we are going to take the no of ratings given for each place.

User	Place	Rating
User A	Place 1	10
User B	Place 1	8
User C	Place 2	8
User D	Place 2	7
User E	Place 1	8
User F	Place 1	7
User G	Place 1	10

Table 3: Popularity-based Recommender

Place	Rating Count
Place 1	5
Place 2	2

The place which has more ratings is more popular. Here place 1 has more ratings compared to place 2, so place 1 is taken as most popular. Hence, we make Popularity-based recommendation system as place 1 for the users over place 2[25-26]. The logic is very simple. Based on the crowd in the place, it is chosen as more popular. SO, based on popularity place 1 is recommended than place 2. Some of the facts on popularity-based recommenders are listed below.

- Rely on purchase history data
- Are often used by online news site

One drawback of popularity-based Recommender systems is that they cannot produce personalized results, this is because they do not consider the user data into account.

3. CORRELATION-BASED RECOMMENDER SYSTEMS

This uses Pearson's r correlation to recommend an item that is most like the item a user has already chosen [27-30]. These correlation-based recommender systems use item-based similarity, like how correlated two items are based on users' ratings. Pearson correlation Coefficient is denoting as r.

- r=1 → Strong +ve Linear relationship among two variables
- r=0 → not linearly correlated
- r=-1 → Strong -ve Linear relationship among two variables

Correlation-based Recommender Systems recommend items based on how well it correlates with other items with respect to user ratings [31-34].

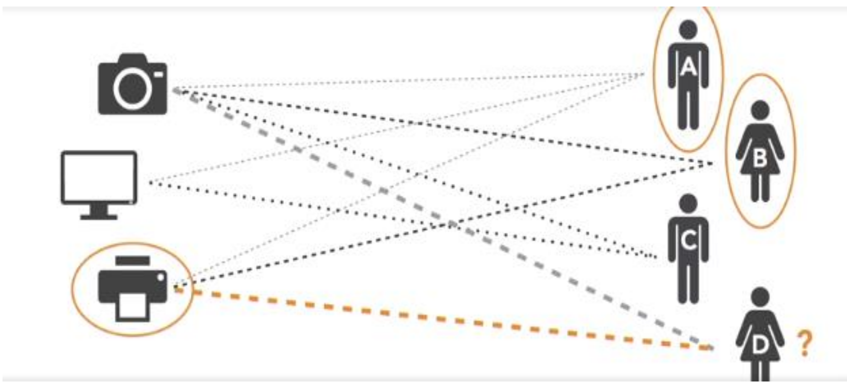

Figure 2: Correlation-based Recommender Systems

User D has reviewed and given 4 stars to camera. Let us observe who else reviewed camera. User A, B and C also reviewed camera. Taking a closer look, look at the ratings each user has given. User A has given 4 stars, user B has given 4 stars and user C has given 2.5 stars. Based on correlation between user ratings, we say that ratings of user A and B are almost like or highly correlated with that of user D. Let us take one more example where in user A and B liked the item. User A and B have given good rating to the printer. Based on how well user A's and user B's reviewed scores correlate with user D review scores of the camera, we recommend the camera to user D as well.

4. CLASSIFICATION-BASED RECOMMENDER SYSTEMS

Classification-based Collaborative Filtering is powered by Logistic Regression and Naïve based classification [13]. Logistic regression is a very simple ML method using which we can predict the value of a numeric categorical variable based on its relationship with predictor variables [35-36]. One of the awesome features of classification-based recommenders is that they can make personalized recommendations, since these recommenders consider the user attributes, purchase history data and contextual data.

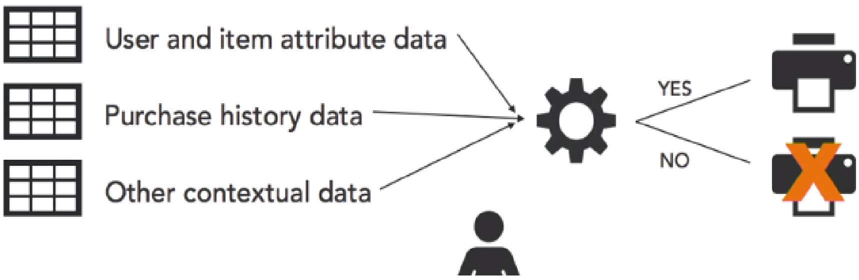

Figure 3: Classification-based Recommender system

For example, if it comes user attributes data, the data set might define key characteristics of user that purchased and did not purchased in the past. For purchase history these data sets describe what purchases users have made and not made in the past [37,38]. We can even think this as a type of transactional historical data. The third type of data it can accommodate is contextual data. By using contextual features, it is possible to break recommendation into unique segments based on things like hours, seasons or user browser history that is tracked by cookies.

Let us take an example of how to make Recommendation system by using logistic regression as a classifier. Consider an example shown in the figure 4. Imagine there is marketing data scientist for a bank. He needs to decide if one of his existing customer or clients is s good candidate to offer the scheme of special term deposit that is currently running by the bank. The representative of the bank definitely will reach out to the client and make him the offer. He uses logistic regression (LR) as a classifier to recommend the offer to the client. Say the client is Mr. John. Mr. John is single, divorced person, who works in management and is not in credit default. He does not have home loans or any personal loans.

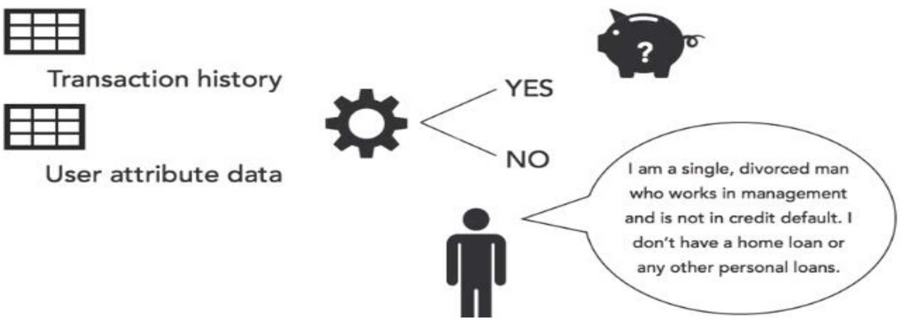

Figure 4: A Logistic Regression Recommender

When we train out logistic regression classifier, we are directed to represent transaction history as well as user attribute data. The history data set of transaction provides very simple attributes on users who have either converted or not converted in the past marketing approach. We also make use of user attribute data set to provide rich description of user who has converted or did not convert in the past. Based on this data, the logistic regression as a classifier will recommend whether to include the client in the marketing initiative. If the model predicts that he will subscribe to the termed deposit upon marketing contract, include the initiative. If the model predicts that he will subscribe, then he should not be contacted with the marketing offer.

5. EVALUATION OF RECOMMENDER SYSTEMS

In order to see how reliable our models are, we have to measure the quality of the predictions that the recommendation systems make [15]. There are several metrics used for the measurement. Precision is measure of model relevancy. It is shown in equation 1.

$$\text{Precision} = \frac{\text{number of items that I liked that were also recommended to me}}{\text{number of items that were recommended}} \quad (1)$$

Precision is the number of items that I liked that were also recommended to me out of the number of items that were recommended. For example if a system recommended 8 items out of four items we liked, then system would have achieved 50 % precision.
Another important metric is recall. Recall is a measure of model's completeness. Recall is shown in equation

$$recall = \frac{\text{number of items that I liked that were also recommended to me}}{\text{number of items that I liked}} \quad (2)$$

Recall is the count of items liked by me that were also recommended to me out of the total items number that I liked. In other words, how completely the recommender system predicts the items I liked? As an example, if the system recommended 8 items out of 10 items I liked, the system would have achieved 80% completion or 80% recall.

6. CONCLUSION

It will always be very difficult and challenging job to make a choice between various choices based on tremendous amount of data available online. Recommendation systems (RS) through online definitely help us to solve these issues. RSs incorporate effective information extraction and filtering processes to do their job competently and accurately. In this article, we have discussed two different types of Collaborative filtering recommendation systems with examples and an insight towards popularity-based recommendation systems. This paper also presents classification-based collaborative filtering using a machine learning algorithm namely Linear Regression. Finally, the different metrics for evaluating the recommendation systems are discussed. The first metric precision is used to measure how relevance the recommendations that were made, the second metric recall is used to measure the completeness of the model, that is, how completely the recommender system predicts the items that are being liked.

REFERENCES

[1] D. Jannach, M. Zanker, A. Felfernig and G. Friedrich, Recommender Systems: An Introduction, New York: Cambridge University Press, 2011.

[2] J. Lu, D. Wu, M. Mao, W. Wang and G. Zhang, "Recommender system application developments: A survey," Decision Support Systems, vol. 74, pp. 12-32, 2015.

[3] E. Rich, "User Modeling via Stereotypes," Cognitive Science: A Multidisciplinary Journal, vol. 3, no. 4, pp. 329-354, 1979.

[4] P. Ghosh and Mahesh T R, "A privacy preserving mutual authentication protocol for RFID based automated toll collection system," 2016 International Conference on ICT in Business Industry & Government (ICTBIG), 2016, pp. 1-5, doi: 10.1109/ICTBIG.2016.7892668.

[5] Pinaki, G., and T. R. Mahesh. "Smart city: Concept and challenges." International Journal on Advances in Engineering Technology and Science 1.1 (2015).

[6] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in Proceedings of the 1994 ACM Conference on Computer supported cooperative work, 1994.

[7] Mahesh, S., Mahesh, T. R., & Vinayababu, M. (2010). USING DATA MINING TECHNIQUES FOR DETECTING TERROR-RELATED ACTIVITIES ON THE WEB. Journal of Theoretical & Applied Information Technology, 16.

[8] Mahesh, T. R., Prabhanjan, S., & Vinayababu, M. (2010). NOISE REDUCTION BY USING FUZZY IMAGE FILTERING. Journal of Theoretical & Applied Information Technology, 15.

[9] Oard DW, Kim J. Implicit feedback for recommender systems. In: Proceedings of 5th DELOS workshop on filtering and collaborative filtering; 1998. p. 31–6.

[10] A A Deshmukh , Ms. Manali Dubal2 , Dr. Mahesh TR , Mr. C R Chauhan, "DATA SECURITY ANALYSIS AND SECURITY EXTENSION FOR SMART CARDS USING JAVA CARD" , International Journal of Advanced Information Technology (IJAIT) Vol. 2, No.2, April 2012

[11] Kumar, V.V., Raghunath, K.M.K., Muthukumaran, V. *et al*. Aspect based sentiment analysis and smart classification in uncertain feedback pool. *Int J Syst Assur Eng Manag* **13,** 252–262 (2022). https://doi.org/10.1007/s13198-021-01379-2

[12] H. K. Shashikala, T. R. Mahesh, V. Vivek, M. G. Sindhu, C. Saravanan and T. Z. Baig, "Early Detection of Spondylosis using

Point-Based Image Processing Techniques," 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 2021, pp. 655-659, doi: 10.1109/RTEICT52294.2021.9573604.

[13] Kuldeep Sharma, Mahesh T R , Big Data Technology for Developing Learning Resources, Journal of Physics: Conference Series IOP Publishing Ltd, May 2021 doi:10.1088/1742-6596/1979/1/012019

[14] Dhiman, G., and V. Kumar. "Harmony Search and Nature Inspired Optimization Algorithms." (2019): 857.

[15] Gowramma, G. S., T. R. Mahesh, and G. Gowda. "An automatic system for IVF data classification by utilizing multilayer perceptron algorithm." ICCTEST-2017 2 (2017): 667-672.

[16] Dharahas, R. T., & Mahesh, T. R. (2021). A Pragmatic Approach for Detecting Brain Tumors Using Machine Learning Algorithms. BIOSCIENCE BIOTECHNOLOGY RESEARCH COMMUNICATIONS Special Issue, 14(11).

[17] McSherry D. Explaining the pros and cons of conclusions in CBR. In: Calero PAG, Funk P, editors. Proceedings of the European conference on case-based reasoning (ECCBR-04). Madrid (Spain): Springer; 2004. p. 317–30.

[18] T. R. Mahesh, V. Dhilip Kumar, V. Vinoth Kumar, Junaid Asghar, Oana Geman, G. Arulkumaran, N. Arun, "AdaBoost Ensemble Methods Using K-Fold Cross Validation for Survivability with the Early Detection of Heart Disease", Computational Intelligence and Neuroscience, vol. 2022, Article ID 9005278, 11 pages, 2022. https://doi.org/10.1155/2022/9005278

[19] Karthick Raghunath K. M, V. Vinoth Kumar, V.Muthukumaran, krishna kant singh, Mahesh T R, Akansha Singh, "DETECTION AND CLASSIFICATION OF CYBER ATTACKS USING XGBOOST REGRESSION AND INCEPTION V4", Journal of Web Engineering, RIVER PUBLISHERS, Vol 21, Issue 4, 2022 https://doi.org/10.13052/jwe1540-9589.21413

[20] Kumar, K. Vinoth, and S. Bhavani. "Trust Based Multipath Authentication Protocol for Mobile Ad-Hoc Network." *Journal of Computational and Theoretical Nanoscience* 12.12 (2015): 5479-5485.

[21] P. Chaitanya Reddy, R. M. S. Chandra, P. Vadiraj, M. Ayyappa Reddy, T. R. Mahesh and G. Sindhu Madhuri, "Detection of Plant Leaf-based Diseases Using Machine Learning Approach," 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), 2021, pp. 1-4, doi: 10.1109/CSITSS54238.2021.9683020.

[22] Adomavicius G, Zhang J. Impact of data characteristics on recommender systems performance. ACM Trans Manage Inform Syst 2012;3(1).

[23] R. Pasumarty, R. Praveen and M. T. R, "The Future of AI-enabled servers in the cloud- A Survey," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2021, pp. 578-583, doi: 10.1109/I-SMAC52330.2021.9640925.

[24] A. S. Harpale andY.Yang, ``Personalized active learning for collaborative _ltering,'' in Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2008, pp. 91_98.

[25] A. S, R. F. Johnson, R. k. N, M. T R and V. V, "Predicting the number of new cases of COVID-19 in India using Survival Analysis and LSTM," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2021, pp. 1-4, doi: 10.1109/I-SMAC52330.2021.9640899.

[26] L. Li, D. Wang, T. Li, D. Knox, and B. Padmanabhan, ``SCENE: A scalable two-stage personalized news recommendation system,'' in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. (SIGIR), 2011, pp. 125_134.

[27] M. R. Sarveshvar, A. Gogoi, A. K. Chaubey, S. Rohit and T. R. Mahesh, "Performance of different Machine Learning Techniques for the Prediction of Heart Diseases," 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), 2021, pp. 1-4, doi: 10.1109/FABS52071.2021.9702566.

[28] D. Zeleník and M. Bieliková, ``News recommending based on text similarity and user behaviour,'' in Proc. WEBIST, 2011, pp. 302_307.

[29] Mahesh T R, Dr. Mohan Kumar Naik , A Comprehensive Review of Behavioral Customer Segmentation For A Better Understanding, International Journal of Computer Science and Engineering, Volume 8 Issue 1, 1-4, January 2021

[30] S. Bourke, "The Application of Recommender Systems in a Multi Site, Multi Domain Environment," in Proceedings of the 9th ACM Conference on Recommender Systems, 2015

[31] M. Narayanan and A. K. Cherukuri, "A study and analysis of recommendation systems for location-based social network (LBSN) with big data," IIMB Management Review, vol. 28, pp. 25-30, 2016.

[32] Mahesh T R, Dr. Mohan Kumar Naik , "Analysis of Academic Performance in massive Open Online Courses (Moocs) Using Process mining" International Journal of Computer Trends and Technology, Volume 68 Issue 12, 21-25, December 2020

[33] Gowramma GS, Mahesh TR, Gowda G (2017) An automatic system for IVF data classification by utilizing multilayer perceptron algorithm. In: ICCTEST-2017. ISBN 978-81-931119-5-6, vol 2, pp 667–672. https://doi.org/10.21647/icctest/2017/49043

[34] Asha S C, T R Mahesh, V Vivek and Suresh M B, The Importance of Teacher's Mental Health and Wellness for Quality Learning in Classrooms during COVID-19 Pandemic April 30, 2021, available at SSRN: http://ssrn.com/abstract=3837304

[35] S. Vairachilai, M. Kavithadevi and M. Raja, "Alleviating the Cold Start Problem in Recommender Systems Based on Modularity Maximization Community Detection Algorithm," Circuits and Systems, vol. 7, pp. 1268-1279, 2016.

[36] K. Chenchireddy, V. Kumar and K. R. Sreejyothi, "Investigation of Performance Vector Control Single-Phase Induction Motor," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 887-891, doi: 10.1109/ICACCS51430.2021.9441773.

[37] Kuldeep Sharma, Mahesh T R , Big Data Technology for Developing Learning Resources, Journal of Physics: Conference Series IOP Publishing Ltd, May 2021 doi:10.1088/1742-6596/1979/1/012019

[38] Basheer, S., Anbarasi, M., Sakshi, D.G. *et al.* Efficient text summarization method for blind people using text mining techniques. *Int J Speech Technol* **23,** 713–725 (2020). https://doi.org/10.1007/s10772-020-09712-z

BIOGRAPHIES OF AUTHORS

	T. R. Mahesh is serving as Associate Professor and Program Head in the Department of Computer Science and Engineering at Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, India. Dr. Mahesh has to his credit more than 40 research papers in Scopus and SCIE indexed journals of high repute. He has been the editor for books on emerging and new age technologies with publishers like Springer, IGI Global, Wiley etc. Dr. Mahesh has served as reviewer and technical committee member for multiple conferences and journals of high reputation. His research areas include image processing, machine learning, Deep Learning, Artificial Intelligence, IoT and Data Science.
	V. Vivek is having 13 years of experience in teaching and research domains. His area of expertise includes Distributed Systems, Cloud Computing, Computer Networks, Agent-based Computing. He has completed his Master's and Doctoral degrees from the School of Computer Science and Technology, Karunya University, INDIA. He has received the CISCO certification in CCNA (Cisco Certified Network Associate) from CISCO Systems. He is also a part of the CSICO Networking Academy team as a CISCO Certified Instructor for more than eight years. He has published research articles in leading journals (SCI, and Scopus) and was a resource person for various guest lectures. He is also actively involved in signing academic MoU's with Infosys, CISCO, and Microsoft and organized various technical FDP's, workshops and hands-on training. Recently has been elevated as IEEE senior member by IEEE. He has worked for universities like Karunya University – Coimbatore and Alliance University- Bangalore and currently working as an Assistant Professor and Program Coordinator in the Department of Computer Science and Engineering (AI & ML) at JAIN (Deemed to be University), Bengaluru.
	**V. Vinoth Kumar** is an Associate Professor at Department of Computer Science, JAIN (Deemed-to-be University), Bangalore, India. His current research interests include Wireless Networks, Internet of Things, machine learning and Big Data Applications. He is the author/co-author of papers in international journals and conferences including SCI indexed papers. He has published as over than 35 papers in IEEE Access, Springer, Elsevier, IGI Global, Emerald etc.. He is the Associate Editor of International Journal of e-Collaboration (IJeC), International Journal of Pervasive Computing and Communications (IJPCC) and Editorial member of various journals.