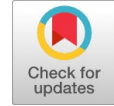# Machine Learning Based Product Comparison for E-Commerce Websites

**Mahalakshmi K, Keerthika R, Sruthi R**

*Abstract: Online shopping through e-commerce has gained widespread popularity among consumers, revolutionizing the operations of businesses in the global market. This paper examines the benefits of e-commerce, such as its convenience and the ease of comparing prices and products, as well as the difficulty customers may encounter when selecting the optimal product. To overcome this difficulty, the paper suggests a real-time online consumer behavior prediction system that anticipates a visitor's purchasing intent using session and visitor data and assesses the effectiveness of Continuous Learning with the Naive Bayes strategy. The article also focuses on developing a recommendation system that strikes a balance between increasing precision and safeguarding users' privacy, utilizing the Prize dataset to assess the system's accuracy. Additionally, the paper delves into the domain of opinion mining, outlining its objectives and responsibilities, such as anticipating sentiment, summarizing aspect-based sentiment, and predicting the helpfulness of online feedback and reviews.*

## I. INTRODUCTION

The use of a collaborative filtering-based recommender system has both positive and negative implications. While it can aggregate and organize the preferences of multiple users to provide relevant recommendations, thus improving user experience and increasing revenue for a website, it can also lead to the potential leakage of users' private information. This article focuses on developing, testing, and validating a recommender system that includes security measures to address this issue. To evaluate the accuracy of the system, the Prize dataset is used, which also influences the algorithm selection. The article aims to strike a balance between improving accuracy and protecting users' privacy, which is a part of an ethical circle that fosters more significant collaboration and improves precision and protection simultaneously, thereby reducing user concerns related to sharing their private data.

**Mahalakshmi K,** Research Supervisor, Department of Computer Science and Engineering, Kalaignar Karunanidhi Institute of Technology, Coimbatore (Tamil Nadu), India. ORCID ID: https://orcid.org/0000-0003-3627-6220

**Prof. Keerthika R,** Department of Computer Science and Engineering, Kalaignar Karunanidhi Institute of Technology, Coimbatore (Tamil Nadu), India.

**Sruthi R\*,** Student, Department of Computer Science and Engineering, Kalaignar Karunanidhi Institute of Technology, Coimbatore (Tamil Nadu), India. E-mail: sruthilakshmi2000@gmail.com, ORCID ID: https://orcid.org/0009-0001-4651-765X

### A. Opinion Mining

Item assessment mining is a cycle that follows people's overall attitudes toward a single item. Conclusions are important when deciding on a course of action or selecting amongst several options. The goal of data collection has always been to find out what other people believe. The availability of conclusion-rich assets, for example, internet audit locations and particular websites, and challenges grow in comprehending the evaluations of others.

### B. Tasks in Opinion Mining

Opinion analysis aims to predict the sentiment of a given piece of writing, whether it is positive or negative. Unfortunately, due to its relatively low demand, job opportunities in the field of product opinion analysis tend to be overlooked. The tasks involved in opinion analysis include detecting subjectivity, predicting sentiment, summarizing aspect-based sentiment, summarizing contrastive viewpoints, and summarizing text for opinions. Additionally, it involves predicting the helpfulness of online comments and reviews.

## II. LITERATURE SURVEY

1) According to M.Rejc (2011), in a free domain, machine administrators are compelled to acquire positive auxiliary administrations, which may include reimbursement for dynamic vitality disasters. The compensation typically involves buying high-quality goods over a long period and supplementing them with additional short-term high-quality purchases to account for regular fluctuations. Brief timeframe period power purchases necessitate an accurate and quick brief timeframe period measuring strategy that must be successfully relevant in day ahead of time shows. This study proposes a fresh brief timeframe period dynamic quality loss gauge strategy for the specified day using impact stream assessment.[6] This includes momentary load and period gauges, as well as organization geographic estimations, which may be used for quality stream calculations and subsequent dynamic power misfortune counts. A fluffy weight collection of the unusual transitory load and innovation conjecture findings is presented to reduce estimation errors. A novel approach to preprocessing input records has been developed, which considers the methodologies for training the predictive models when performing rough clustering. By utilizing real data from the ENTSO-E connection and conducting tests on the Slovenian power system, the effectiveness of the suggested approach was confirmed. The significant results prove the improved precision of the proposed strategy.

Retrieval Number: 100.1/ijitee.F95610512623
DOI: 10.35940/ijitee.F9561.0512623
Journal Website: www.ijitee.org

38

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

2) In his 2015 work, F. Sulla explained how this paper applies multiple direct regression methods (MLR) to analyze the significant correlations between the damping of electromechanical modes and the operating conditions of a system, as well as predict future damping values based on current market estimates for power generation and consumption. The study utilizes data from the Nordic power system and initially develops a static MLR model to understand the variability of damping.[3] If day-already commercial center forecasts are correct, test results show that the suggested processes can effectively predict around 90 percent of the low damped working conditions uncovered throughout the year. These findings imply that the approaches might be used to provide early warnings about future working situations with minimal damping.

3) Because of the discontinuous concept of wind, Zhi Li (2016) proposed that large-scale joining of wind time brings enormous asking conditions to the consistent activity of power frameworks. The variation in breeze innovation has a significant impact on unit duty. Effective decision-making regarding wind power is crucial in addressing the challenges posed by energy system operation in uncertain circumstances, in a practical and specialized manner.[1] This study suggests a method for predicting wind power in the short-term by using a combination of an Extreme Learning Machine (ELM) and a model for adjusting errors. To begin, an ELM is used to assess the transitory wind quality. The super brief timeframe period wind quality determining is therefore obtained essentially based on handling the short timeframe period expecting mistakes by determination procedure.

4) In 2013, L. Xie conducted a study that presents valuable insights into utilizing extremely large data sets for power system operations, control, and security, which are challenging to analyze using conventional database tools and commonly referred to as big data. The paper focuses on three main aspects: identifying key features from the data, integrating it into power system applications, and providing examples of its conventional usage in the utility industry.[4] The following examination tasks based on large measurements approach are described: restorative, prescient, managed, and adaptable. The study also follows a few investigation themes relating to resource executives, activity planning, continual monitoring, and issue location/security that provide further prospects but require further investigation.

5) Azadeh (2008) suggested a combined approach that employs fuzzy logic, data mining, and time series analysis to forecast and evaluate irregular and monthly variations in energy demand, particularly in countries like China and Iran where data records may not be fixed. Since using traditional fuzzy logic or time series analysis can be challenging in adjusting uncertain energy demand, this method can be a useful alternative. However, a standard basis is required to develop fuzzy frameworks, which may not be readily available.[8] Therefore, the technique of removing the standard base using separation rule methods, known as FLT, is used when required. Additionally, a reality mining method, known as FDM, is used to separate the general guideline basis. The most favored approach for collecting time-related data involves choosing between linear (ARMA) and non-linear models, and then assessing for nonlinearity using the McLeod-Li test. If the data is found to exhibit nonlinearity, a suitable non-linear model is selected and compared with the preferred ARMA model. Finally, either the non-linear or the ARMA model is chosen as the time collection variation, based on the comparison results.

6) DX Niu (2012) introduces a model for short-term load prediction, called Bayesian neural network (BNN), which is trained using hybrid Monte Carlo (HMC) computations. The weight vector limit of the Bayesian neural network is a multidimensional variable that is irregular. While learning, the Bayesian neural network is viewed as a Hamiltonian dynamic structure, with the working variable being the load vector.[5] The HMC rule set is used to control the weight vector limit separately for normal pre-distribution and Cauchy pre-transmission. The BNN model is evaluated by comparing it with the ANN model discovered by BP calculations, as well as the Bayesian neural networks discovered by Laplace and HMC calculations. The models are evaluated based on the 25- day hourly load predictions of four different seasons: Spring, Summer, Autumn, and Winter. The overall performance of the models is evaluated using the Mean Squared Error (RMSE) and Mean Total Index Error (MAPE).

7) According to Y. Wang (2011), the development of intensity devices and activities is driven by the need for short term exposure measurement (STLF). Given the stress caused by China's rapid construction, this article proposes a unique two-step approach for a 50/50 estimate. In the first step, time management techniques are used to forecast the daily workload during the baseline phase.[7] In the second step, the impact of relative variables is assessed to determine the deviation from the timing strategy, which is then incorporated into the results of the first step. What sets this article apart from conventional methods is its comprehensive analysis of the impact of relative variables on the difference between actual workload and expected outcomes of traditional time management approaches.

8) KB Melody et al. (2006) suggest that load metering is essential for a stable and effective energy market. Several pricing strategies have been studied, such as simple iterations, stochastic periods, Kalman filters, computational intelligence, and neural network structures. The proposed method in this paper involves combining the irregular direct recursion method, general exponential smoothing method, and temperature sensitivity analysis to calculate cross-load.[9] The direct recursion method is used to capture lower load requirements on weekends and Mondays than on other weekdays. The general exponential smoothing method is used to estimate non weekend daily loads. Additionally, the temperature sensitivity is incorporated to enhance the accuracy of the weight determination and to account for the effect of temperature on load. The proposed method has been tested and shown to improve the accuracy of load determination in 1996.

9) As per J. Liu's (2015) research, the erratic and unpredictable lifespan of solar photovoltaic (PV) panels has posed a challenge for machine operators looking to effectively utilize PV power plants. Previous models that considered temperature, humidity, and wind speed to determine the optimal age requirement for daily PV power quality were found to be consistently inadequate under extreme weather conditions. Liu also notes that the" AI" air file, which addresses environmental particle problems, has a direct impact on the attenuation of solar radiation and can affect the quality of power produced by photovoltaic panels.[2] To address these challenges, this paper proposes a novel model for PV power measurement that leverages artificial intelligence as a new frontier for information technology. The aim of this project is to study customer behavior using e-commerce mining techniques to study customer behavior and recommend a product link.

10) The field of e-commerce is rapidly expanding, resulting in the generation of vast amounts of data, including transaction logs and details. This data can be utilized to perform accurate market analysis and forecasting. Numerous research efforts have been dedicated to determining the amount of information and its potential to enhance e-commerce opportunities. However, due to the complexity of these data sets, examples are often derived from imitation, leading to time-consuming and intricate searches. To address this issue, several constraints have been implemented, including the gap constraint, which limits the distance between two consecutive replacements of discovered patterns, the rule that requires the rapid updating of patterns to reflect current practices, and the prudence requirement, which establishes reasonable time intervals for the detection of patterns. A measure of the complexity of the changes needed to modify data in web server log documents is presented in the SOM contribution.

## III. EXISTING SYSTEM

The major goal of our research is to investigate consumer behavior utilizing e-commerce mining techniques to mine customer behavior and propose product links. E-commerce is quickly expanding, and these web-portals create massive volumes of data. This data comprises numerous logs and transaction details, and market evaluations and projections may be done quite correctly with this information. Various research attempts have been made to discover the quantum of information and how it might be used to expand e-commerce potential. Because of circles, instances might occur frequently in a follow, resulting in large time expenditure and ridiculous complexity of the inquiry. The hole limitation places a limit on the separation of two continuous exchanges of discovering patterns, the rule essential creates examples to quickly alter the most recent practices, and the conservativeness requirement ensures reasonable time ranges for the discovered patterns. The measure fine-tunes the adjustments required to shift the information stored in the Web Servers Log records to a contribution of SOM.

## IV. PROPOSED SYSTEM

Web mining refers to the process of extracting valuable insights and knowledge from the vast amount of data available on the World Wide Web. It involves the use of various data mining techniques, such as clustering, classification, association rule mining, and pattern recognition, to analyze and discover useful information from web-based data. Web mining has become an essential tool in various fields, such as e-commerce, business intelligence, marketing, and scientific research. By extracting information from web pages, online databases, and social media platforms, The process of web mining has the potential to yield significant insights into customer behavior, market trends, and user preferences. The vast amount of data available on the web can be harnessed effectively through the use of web mining, which is a potent tool for obtaining valuable insights and knowledge. With the increasing importance of online data, web mining will continue to play a critical role in various fields and applications in the years to come. It extracts user product views and purchase data from netflix using automated methods. Web content mining is the process of extracting information from inside a page. This technology provides more sophisticated possibilities than current CRM based solutions. This method will display product pricing from several merchants to show you where to buy the goods at a reasonable price, as well as reviews and availability. K-Nearest Neighbor (KNN) is a supervised learning algorithm frequently utilized to classify data according to how its neighbors are classified. Although it can be used for both regression and classification, KNN is primarily utilized for classification problems. KNN stores all the accessible cases and categorizes new cases based on their similarity to previous cases. Suppose an image of a shape resembling both a rectangle and a triangle, but identification of its type as rectangle or triangle is required. The KNN algorithm can be utilized for this purpose, which relies on similarity measures. The KNN model will explore the new dataset for features similar to those of rectangle and triangle images. The shape will be classified as either a rectangle or a triangle based on the most similar characteristics.
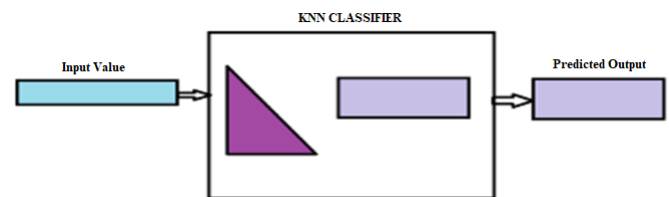


**KNN Classifier**

K- Continuous learning is a machine learning approach that involves continuously training a model with new data to enhance its precision and forecasting capability. Naive Bayes is a widely used classification algorithm that can be employed for continuous learning purposes. The reason for requiring continual learning in machine learning is that various real-world scenarios entail constantly changing and evolving data. With time, the fundamental data distribution may alter, which can lead to a decline in the performance of machine learning models that were originally trained on past data.

In the Naive Bayes algorithm, the likelihood of a sample belonging to a particular class is computed using Bayes' theorem, which considers the presence of its features. By continuously learning with Naive Bayes, the model can incorporate new information and update its probability estimates as new data becomes available. As additional data is integrated, the Naive Bayes model adjusts its probability estimates to incorporate the new knowledge. This feature allows the model to adjust to variations in the data distribution and improve its precision gradually. Continuous learning with Naive Bayes is particularly advantageous in scenarios where the data is continually changing, and the underlying distribution may shift over time. In conclusion, continuous learning with Naive Bayes has the potential to improve the accuracy and resilience of classification models in dynamic and evolving data environments.

## A. Research Methodology

The proposed framework assumes that clothing recommendations will be tailored to an individual client. To achieve this, the framework utilizes KNN computation to identify the K closest neighbors and anticipate the client's preferences.

Reviews are pre-processed and analyzed for sentiment, resulting in either positive or negative feedback. Based on this feedback and other input, the framework recommends the best clothing options.

During the co-positioning process, evaluation targets/words are removed to ensure objectivity. Objective competitors are assumed to be all things/thing phrases, while potential emotion words are descriptors/action words.

The data provided may consist of messages, images, or archives and may be analyzed for subject and topic-specific opinions. The technique of presumptive link-relevance recognition is used to measure word arrangement.

A word-based arrangement model is used for performing monolingual word arrangement, which is commonly employed in tasks such as collocation extraction and label suggestion.

The Continuous Learning with Naive Bayes (CLNB) approach involves training a basic classifier using training data, followed by the creation of subsequent classifiers that focus on instances missed by the previous classifier, until a specified limit is reached.

## B. Preprocessing

The word-based arrangement model utilized by the module is widely employed in tasks such as collocation extraction and label suggestion for monolingual word arrangement. To match a thing/thing stage with its modifiers in phrases and account for potential sentiment words, we use a bilingual word arrangement computation in a monolingual context. However, when applying the traditional arrangement model to our assignment, evaluation target applicants (thing/thing phrases) may align with insignificant words instead of prospective conclusion words like related terms and conjunctions, which are descriptors/action words.
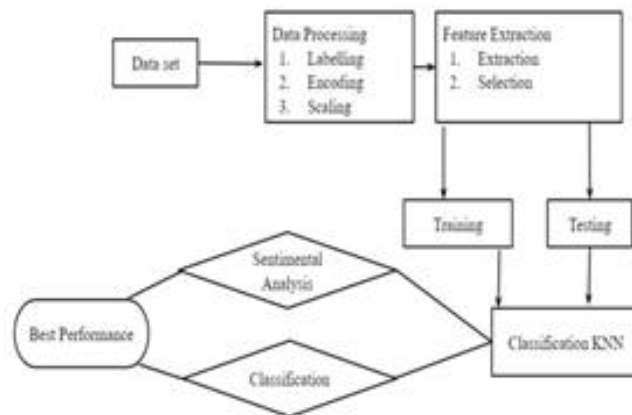


**Architecture Diagram**

## C. Parameter Estimation with CLNB

The layouts created by CLNB Learning should be as stable as possible with median layouts marked. Obtaining all potential offers from the observed information. This shows that standard word formation in preparation for calculations is tedious and illogical. The CLNB learning algorithm is an ideal solution for expediting the training process to address this issue. There is a place on" Neighbor Systems" where you can search for the perfect system; the current agreement, where" neighborhood agreements" are the agreements that could be created by the current agreement.

## D. Acquiring Partial Alignment Links

Various classifiers are employed by this module to evaluate each review, and the outcomes are subsequently combined using techniques like averaging or majority voting. The distinctiveness of each example in the training dataset enables each classifier to offer a unique perspective on the matter being studied. Nevertheless, this method can be unfeasible and time-intensive for certain domains. As noted earlier, although existing syntactic parsing tools may not be capable of precisely capturing the complete syntactic structure of informal sentences, they can still effectively extract certain direct or straightforward syntactic relations. In order to guarantee the accuracy of the employed syntactic patterns, we confine them to immediate dependency relations specified within the reviews.

## E. Audit Analysis and Recommendation

This module identifies a more extensive scope of review involving several leaders, and these long-range vertices are likely to gather a significant amount of data from their neighboring vertices and have a considerable impact on multiple vertices as they take unpredictable walks. A peak is more likely to be reached by a hiker if it is linked to a peak of high magnitude. The most prevalent technique for data collection is positive and negative survey layout. The common procedure is called Majority Voting (MV) and is described by many "experts"; which defines the extremum of a set by considering each classifier's score to be equally significant and determines the last extremum by choosing the most common score prediction from resistance estimates.

Framework provides a multi-area presumption arrangement technique aimed for the community.

Approach can develop exact idea classifiers for several regions while working in a group and dealing with the issue of insufficient marked information by using the thought relatedness across distinct spaces. The sentiment classifier for each region is split into two components: a global component and a region specific component. The global model captures the overall bias information that is common across multiple regions, while the region-specific models capture the specific assessment expressions of each region. Propose that space explicit CLNB feeling information be removed from both identified and unlabeled instances and used to improve the learning of area explicit notion classifiers.

Furthermore, suggest using the previously broad idea information all in all reason assumption vocabularies to steer the learning of the global notion classifier. Furthermore, suggest combining similarities between different spaces into an approach as regularization over the area explicit conclusion classifiers to accelerate the exchange of assumption data between comparable regions. Convert the approach model into a curved improvement issue. Also, to associate a sped-up calculation with successfully explaining the model of our technique and offer an equivalent calculation to further increase its efficacy when regions to be broken down are massive. Exploratory results on benchmark datasets demonstrate that the strategy can successfully improve the presentation of multi-space assessment characterization and outperform current techniques.

## V. OBTAINING USER/INTERNAL INFORMATION

Data quality should be checked before using machine learning or data mining algorithms. • This module collects the collected data of the user and the item. • The report in which the user just saw the product/purchased the product/identified both.

### A. Results of Naive Bayes Continuous Learning Experiments

Collaborative filtering algorithms generate recommendations for customers by discovering direct relationships between customers or between products, such as similar products bought by customers themselves, or by discovering implicit associations between customers and products based on their interactions, such as using dimensionality reduction techniques. Collaborative filtering is a widely used approach for making recommendations by analyzing customer interactions with products, such as finding relationships between customers or products. However, a significant drawback is the lack of data on customers who have not interacted with any products. On the other hand, content-based algorithms are not affected by this problem but may not perform as well as collaborative filtering. They rely on structured and consistent product metadata. Understanding The digital experience is personalized for each customer by utilizing data.

### B. Result of Experiments

For the main study, the information index was randomly divided into preparation and test sets without taking trade

requests into consideration. Thus, the analysis focused solely on promotions and prices and excluded the most recent event. Although dividing the dataset by time would provide a more meaningful score and allow for measuring the impact of recency, there were few buying opportunities in the last few months, with only 0.07 percent of the last 10 percent of total deals being purchased. The data set was then split in time, with the top 80 percent of results modeled in a sequential query, and predictions made for the bottom 20 percent. During this period, only 0.6 percent of all bargains were purchased.

**Table- I: Result Analysis**

| MODEL | AUC | MAP CLICKS | MAP WANTS | MAP PURCHASES |
|---|---|---|---|---|
| CLNB | 0.751516 | 0.014772 | 0.011147 | 0.009899 |
| KNN | 0.735779 | 0.004435 | 0.009703 | 0.006054 |
| BPR | 0.728643 | 0.012015 | 0.005894 | 0.013378 |

## VI. CONCLUSION

The framework provides a cross-domain, community oriented approach to guesswork. The proposed method can effectively train concept classifiers for multiple domains in a consistent manner and overcome the issue of insufficient labeled data by leveraging the concept affinity across various domains. Each domain's sentiment classifier comprises two components, namely global and domain-specific segments. The global model can capture general slope information shared by different spaces, and explicit surface models are used to capture individual assessment joints for each space. Propose to remove spatially explicit CLNB sensation information from labeled and unlabeled examples and use it to improve classifier learning of explicit domain concepts. Furthermore, we propose to use the foregoing general information about the concepts of all conjecture vocabularies to guide the training of the global concept classifier.

## VII. FUTURE WORK

Also suggest combining similarities between different spaces in an approach like a regularization in the field of explicit inferential classifiers to stimulate the exchange of hypothesis data between comparable fields. Draw a model approach to the curve improvement problem. Also introduce the accelerated calculation to an efficient model to explain our methodology and propose an equivalent calculation to further improve its efficiency when the areas to be parceled are monstrous. Results from the exploration of comparative data sets show that this approach can successfully improve the representation of multi-spatial assessment features and significantly outperform standard techniques.

# Machine Learning Based Product Comparison for E-Commerce Websites

## DECLARATION

| | |
|---|---|
| Funding/ Grants/ Financial Support | No, I did not receive. |
| Conflicts of Interest/ Competing Interests | No conflicts of interest to the best of our knowledge. |
| Ethical Approval and Consent to Participate | No, the article does not require ethical approval and consent to participate with evidence. |
| Availability of Data and Material/ Data Access Statement | Not relevant. |
| Authors Contributions | Mahalakshmi K developed the study's concept and design, carried out data collection and analysis, and composed the initial manuscript. Keerthika R provided valuable input to the study design, contributed to data analysis and interpretation, and made significant revisions to the manuscript. Sruthi R was involved in data collection and analysis, and provided critical feedback and edits to the manuscript. All authors have carefully reviewed and approved the final version of the manuscript. |

## REFERENCES

1. Z. Li et al., "Short-term wind power forecast with error correction," Protection Control Modern Power Syst., vol. 1, no. 1, pp. 1-8, 2016.
2. J. Liu, W. Fang, X. Zhang, and C. Yang, "An improved photovoltaic power forecasting model with the assistance of aerosol index data," IEEE Trans.Sustainable Energy, vol. 6, no. 2, pp. 434–442, Apr. 2015. [CrossRef]
3. F. Sulla, M. Koivisto, and J. Seppanen, "Statistical study and forecasting of damping in the Nordic power system," IEEE Transactions on Power Systems, vol. 30, no. 1, Jan. 2015, pp. 306-315. [CrossRef]
4. M. Kezunovic, L. Xie, and S. Grijalva, "The role of big data in improving power system operation and protection," in Proc. IEEEIREP Symp.Rethymnon Bulk Power Syst. Dyn. Control-IX Optim.Security ControlEmerging Power Grid, 2013, pp. 1–9. [CrossRef]
5. D. X. Niu, H. F. Shi, and D. D. Wu, "Short-term load forecasting using Bayesian neural networks learned by hybrid Monte Carlo algorithm,"Appl. Soft Comput., vol. 12, no. 6, pp. 1822–1827, 2012. [CrossRef]
6. M. Rejc and M. Panto, "Short-term transmission-loss forecast for the Slovenian transmission power system based on a fuzzy-logic decision technique," IEEE Trans. Power Syst., vol. 26, no. 3, March 2011, pp. 1511-1521. [CrossRef]
7. Y.Wang, Q. Xia, and C. Kang, "Secondary forecasting based on deviation analysis for short-term load forecasting," IEEE Trans. Power Syst., vol. 26, no. 2, pp. 500–507, May 2011. [CrossRef]
8. A. Azadeh, M. Saberi, S. F. Ghaderi, and V. Ebrahimipour, "Improved estimate of power demand function by the integration of fuzzy system and data mining technique," Energy Convers.Manage., vol. 49, no. 8, pp. 2165-2177, 2008. [CrossRef]
9. K. B. Song et al., "Hybrid load forecasting method with analysis of temperature sensitivities," IEEE Trans. Power Syst., vol. 21, no. 2, pp. 869–876, May 2006. [CrossRef]

## AUTHORS PROFILE

**K. Mahalakshmi** is a research supervisor in Information and Communication Engineering under Anna University, Chennai. She has five years of experience in Software Industry and sixteen years experience in teaching. She completed her PhD in February 2015. Her Ph.D. is on Software requirement engineering under the Faculty of Information and Communication Engineering, Anna University, Chennai. She did her Master of Technology in Information Technology from AAI, Deemed University, Allahabad. She also did her Master of Business Administration in Human Resources from Annamalai University and Bachelor of Engineering in Computer Science and Engineering from Annamalai University. She is now working as Professor, in The Department CSE, Kalaignarkarunanidhi Institute of Technology, Coimbatore. She has serviced as Career Counselor, Motivator, and Soft Skill Trainer. She is also a reviewer in Anna University Annexure Journals like Springer Journals - Cluster Computing (CLUS), Design Automation for Embedded Systems (DAEM), Wireless Personal Communications, Cognition Technology & Work (CTWO), and Multimedia Tools and Applications. She is a Professional Member of ISTE, CSI, SEEE, IAENG, and ISRD. She has published many articles. Her research and publication interests include Data Mining, Text Classification, Requirement Engineering - Software Engineering, Big Data Analytics & Data Mining, Cyber Security.

**R. Keerthika** is a highly experienced professor with a total experience of 14 years and 6 months in the education management industry. She holds a B. Tech., M. Tech., and Ph.D. degree and has published extensively in her field with 14 articles in international journals, out of which 11 are indexed in Scopus and 2 in Web of Science. She has also presented her research in 6 international conferences and 3 national conferences, and her work has been cited 17 times, indicating the impact of her research. Dr. Keerthika is actively involved in teaching and community service within the scientific community.

**R. Sruthi** is a motivated student pursuing a Master's degree in Computer Science and Engineering with a keen interest in utilizing technology to tackle real-world problems. With a Bachelor's degree in Information Technology already under her belt, she has focused her research on data science and artificial intelligence. Her pursuit of a Master's degree is driven by her desire to further develop her skills and expertise in these areas. Sruthi has gained practical experience through internships with major companies such as Cognizant Technology Solutions and Philips Innovation Campus, which have prepared her for a successful career in the field of technology.