International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

March 2022 | Issue 50 | Volume 14 | Number 1 | Pages 89-94

A PROFIT-DRIVEN STACKING MODEL FOR EFFECTIVE CHURN PREDICTION

J.N. Victor

Department of Computer Science, St. Joseph's College, Bharathidasan University, Trichy, India
drjudenirmal@yahoo.com

Abstract- When an organization's primary focus is on providing customer-based services, employee churn is a huge problem. Customers have a strong inclination to churn and move to a competitor's product or service, which makes operating in highly saturated marketplaces even more stressful. We're looking for churners, but we also want to make sure the company remains profitable. As a result of this research, an MPDS (multi-level profit-driven stacking model) has been developed, which can be used to forecast customers who will leave. As a result of the multi-level model, even the most complex sets of data may be handled with ease. Using real-time churn prediction data, researchers tested the recommended MPDS model. The result found that it performs well, producing its model for usage in real-time surroundings.

Keywords: Prediction of Churn, Stacking, Multi-Level Modeling, Profit-Driven Modeling, Behavior-Based Modeling.

1. INTRODUCTION

Because of the fierce rivalry for clients, customer-based markets are always a step ahead. Online strategies for customer retention and acquisition have played an important role in these markets' recent growth. Many industries, such as telecommunications, have reached saturation point, with over 90% of users tied to a single service provider [1]. The pressure on organizations functioning in saturated marketplaces is enormous, as enticing offers from competing providers might be critical points for clients to quit their existing provider [2]. While resources are devoted to luring in new clients, keeping existing clients is a problem that must be overcome. New customer acquisition is more expensive, according to the research [3]. Instead, then spending money on a possible new consumer, it makes sense to provide better deals to existing customers and increase the likelihood that they will stick around. Customer retention initiatives are also conducted by organizations on a regular basis to help with this [4]. Identifying churners from a large customer base is the primary goal of customer retention program.

Organizations utilize prediction-based modeling tools to effectively identify potential churners [5]. In order to identify likely churners, behavioral and historical data about customers is collected. The majority of existing models use accuracy as a criterion for assessing the churn prediction model's efficacy. For a company, however, profit is critical in determining the performance of models since successful retention efforts must identify consumers who leave because they are not profitable [6].

High-risk consumers would be a waste of time because they are slow to respond [7]. So, advertising to keep clients should be aimed at the right ones [8]. If you want to build a churn prediction model that works, you need to separate out those customers who will not revert from those that will [9]. It is also critical to identify customers at the right time, as waiting a longer period of time increases the likelihood of losing customers. This article suggests a churn prediction model that considers all of the above criteria to produce an effective and profitable business model. The other units are organized as follows: section 2 has a literature analysis of existing churn prediction models; section three contains an in-depth look at the MPDS architecture. The fourth section of the chapter will reveal the discussion. The next chapter will conclude the results to deliver a sufficient and profit-oriented churn forecast model.

2. INTERCONNECTED WORKS

Churn forecasting is an important aspect of every business's operation. As a result, it is being deployed and used in a variety of businesses. This section discusses some of the extant works in the literature. Hoppner, et al. [10], presented a churn forecast for assessment tree-based replica that focuses on raising the levels of organizational benefit in the process rather than predicting churn itself.

This study implements the profit-maximizing determination trees that specify the most beneficial to be churners, and it also contributes to the provision of retention mechanisms by improving the outcomes of retention campaigns, among other things. Profit levels are calculated on the basis of accuracy metrics and other factors. When Verbraken and colleagues published their churn prediction model [11], they included the metric Expected Maximization Profit for Customer Churn (EMPC), which can be used to identify profitable clients.

In addition, the model identifies the most suitable subset of the client floor to mark for the retention movement to be adequate. An Alboukaey, et al. [12] prediction model for mobile telecom churn that takes into account the temporal character of the data during the forecast process has been developed. This model monitors the behavior of customers on a continuous basis in order to detect changes in their behavior. These modifications are widely regarded as the most important markers of customer attrition. It is only by precisely identifying them that retention initiatives can be targeted at churners at the appropriate moment, resulting in enhanced results. There have been several models in the literature that have been developed on the basis of finding behavioral patterns from the input data in order to make good churn predictions.

Vafeiadis, et al. [13] developed an aggregation-based model, whereas Castanedo, et al. [14] developed a period-based model, and Hung, et al. [15] generated a churn prediction standard to anticipate client churn. Ozmen, et al. [16] provided a model for predicting churn in the telecom industry that makes use of ant colony optimization. An ant colony optimization model for churn prediction is used in this work to develop a multi-objective technique that is cost-sensitive and created using an ant colony optimization model. This model performs cost-based processing and takes into account a number of different process objectives. Given the complexity of neural networks and their ability to operate on data depending on user behavioral patterns, they have been applied in a variety of fields.

In addition to the neuro-fuzzy model developed by Abbasimehr, et al. [17], other churn prediction models that have been developed include the comparison model developed by Vafeiadis, et al. [18]. The Clement, et al. [19] and the Omar, et al. [20]. As deep learning has gained in popularity, churn prediction models have been developed for use in the deep learning architecture as well as other architectures. Cenggoro, et al. [21] presented a deep learning architecture for churn prediction that was based on machine learning. In order to make predictions, this model makes use of deep learning as well as vector embedding techniques. The vectors generated by this model were shown to have extremely high discriminating levels between churners and non-churners, indicating that they were highly successful.

Kostic, et al. [22] suggested a theory-based grid instance for churn prediction, and it was based on their findings. The users of the network are represented as graph nodes. As a result, each node represents the customer's entire set of characteristics. They are used to determine the relationship between users, their behavioral characteristics, and their likelihood to churn. Richter, et al. [23] offered a social group-based approach that is similar to the one described above. Wei, et al. [24] established a call-based model for churn prediction that was based on call volume. In order to anticipate churn, this model makes advantage of call information as critical components.

Amin, et al. [25] established a three-phase model for predicting turnover in the financial services industry. In order to determine churn, this research makes use of a feature selection method, a knowledge-based system, and a simulated expert module. The following are examples of machine learning models-based churn prediction models: an analysis-based model developed by Kavitha, et al. [26] and a Logistic Regression-based model developed by Jain, et al. [27]. Churn prediction public Telecom Dataset is provided in [28]. Using heterogeneous ensemble stacking and minority upliftment, Sivasankar, et al. [29] proposes a churn prediction model that is more effective than the traditional method. An intelligent enterprise's guide to big data analytics is discussed in [30] It shows that this notion is a set of data analysis that drives and deploys future company strategy. Likewise, the research also provided deep learning as one method to deal with standard data processing technique limits encountered with enormous data. NOx emissions in Iran are analysed using Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) methodologies. The NOx emission prediction using GA, PSO, and ACO linear and nonlinear equations are presented in [31].

3. MULTI-LEVEL PROFIT DRIVEN STACKING (MPDS) MODEL

Churn prediction is a complex process that involves effective identification of a customer's behavior. Customer behavior tend to vary considerably in nature. Hence, effective identification requires models that can handle the complex data to provide improvement in the prediction process. The proposed Multi-Level Profit Driven Stacking (MPDS) model proposes two layers of processing to perform this process. The initial layer performs cost ratio identification for models, and the second layer contains two level stacking model that provides highly effective predictions.

3.1. Preprocessing and Segregation of Data

Typical input data for the churn prediction method includes customer usage levels as well as payment information for each individual customer. Using publicly available churn prediction data [28], we were able to extract this information. The dataset consists of 3,333 records that were gathered from different clients. There are 21 attributes in total in the data set. The churn class attribute has the value true or false, and it is a Boolean attribute. The information is comprised of a blend of categorical and numerical characteristics. The qualities are defined in more significant detail in Table 1.

Table 1. Depiction of quality

Quality and Attribute Model	Types of Data	Attribute Selection Process	Types of Data Identified
Position	Categorical	Number of evening calls	Numerical
Length of Account	Numerical	Number of evening signs	Numerical
Locality code	Numerical	Number of night minutes	Numerical

Phone number	Categorical	Number of night calls	Numerical
Global plan	Categorical	Number of night charge	Numerical
Voice mail plan and process messages	Categorical	Total global minutes and call process	Numerical
Number voice mail messages	Numerical	Total global valid call process	Numerical
Number of day minutes	Numerical	Total global charges	Numerical
Number of day calls	Numerical	Consumer service-based process	Numerical
Number of day charge	Numerical	Mixing	Boolean
Number of evening minutes	Numerical	Churn	Boolean

3.2. Model Based Profit Identification

The initial process in this work is to determine the efficiency of each of the classifier models in terms of the cost levels incurred by them. Most of the churn prediction model are based on providing good performance by providing improved metrics. However, metrics do not correspond to increased profits in all cases. Hence it becomes necessary to consider profit levels along with the metrics, when performing predictions for organizations. Profit levels cannot be directly calculated. They are identified as a function of cost, which is the level of loss incurred due to losing a customer. Cost levels are inversely proportional to profit. Higher cost leads to reduced profit and vice versa. The cost matrix [29] used in this work is exposed in Table 2.

Table 2. Levels of matrix and process of cost [29]

Actual / Predicted	Churn	Not Churn
Churn	$\gamma(c_o + c_a) + (1-\gamma)(CLV + c_a)$ [C_{TP}]	$c_o + c_a$ [C_{FP}]
Not Churn	CLV [C_{FN}]	0 [C_{TN}]

where, γ provides the ratio of customers who were successfully retained, c_o and c_a depicts the cost of the offer and the cost incurred for administrative maintenance. The matrix presents the cost of predicting a True and False Positive and True and False Negative like $(C_{TP}), (C_{FP}), (C_{TN})$ and (C_{FN}).

the final cost for a model i is given by Equation (1) [29].

$$\text{Cos}t_i = y_i(c_i C_{TPi} + (1-c_i)C_{FNi}) + \\ +(1-y_i)(c_i C_{FPi} + (1-c_i)C_{TNi}) \tag{1}$$

Besides c_i and y_i are the predicted and mentioned actual class labels by specify model i, C_{TP}, C_{FP}, C_{TN}, C_{FN} and it was acquired from the price matrix.

Model significance identified the set by the cost levels exhibited by the classifier models. The models are trained and predicted over the training data. Based on the predictions cost for each model is calculated. This cost represents the absolute value. Cost ratio of a model m is determined by Equation (2).

$$CostRatio_m = \frac{Cost_m}{\max_0^i(Cost_x)} \tag{2}$$

This represents the inversed significance of the models. Higher cost ratio denotes lower significance.

3.3. Level 1 Stacking Model Building

The proposed work operates on three heterogeneous models; decision tree model, logistic regression model (LRM) as well as naive bayes classifier (NBC). Decision tree is a tree-based classifier model, while naive bayes is probability based, and logistic regression is rule based models. The variations in models ensure efficiency and better rule building capability of the models. The models are trained and predicted over the training data. The proposed stacking model comprises of two levels; the initial level uses the training data for model building, while the second level model uses predictions from the first level for model building. The training data is passed to the heterogeneous models for model building. The validation data is passed to the trained models and predicted. Results from the validation data are integrated along with the class column depicting the status of churn. The cost ratio obtained from the previous phase is multiplied with the predictions, and the training data for the level 2 phase is obtained.

3.4. Level 2 Model Building and Prediction

The level 2 modeling phase uses logistic regression as the training model. This phase operates by analyzing predictions from the previous model, and does not consider the actual data into the processing. Hence, this level performs a stacking-based operation. This phase requires a model that exhibits rule-based analysis, as the input data is mostly linear in nature. Since the problem domain corresponds to binary classification, logistic regression is selected for processing. The preparation data built from the previous model is used for model training. Final predictions are performed by passing the test data to the first level of models. Predictions are obtained from all the models, and the predictions are multiplied with the corresponding profit levels of each of the models. The obtained set is passed to the second level logistic regression model, which provides suitable predictions.

4. FINDINGS, RESULTS AND DISCUSSION

MPDS measure has been executed using scikit library through Python. This Performance obtained for the MPDS model is presented in Table 3. The MPDS model has been observed to exhibit high performance levels, with overall accuracy of 98%, and low error levels.

Table 3. Performance of MPDS model

Techniques and Formations	Model of MPDS
False Positive	0.01
True Positive	0.93
True Negative	0.99
False Negative	0.07
Evoke	0.93
Accuracy	0.95
Precision	0.98
F-Measure	0.94

Figure 1 shows the Receiver Operating Characteristic arc depicting from churn forecast level of MPDS replica is presented in chart. ROC curve depicts the relationship between the true positive rate, which represents the churn prediction levels, and the false positive rate, which represents the level of non-churners predicted as churners. High levels of positive rate of true value and low levels of positive rate of false value are indicative of effective classifier models. The node of the graph is located on the top left part, which indicates that the model depicts high TPR and low FPR levels. This indicates that the MPDS model exhibits very low error levels and very high churn prediction level.

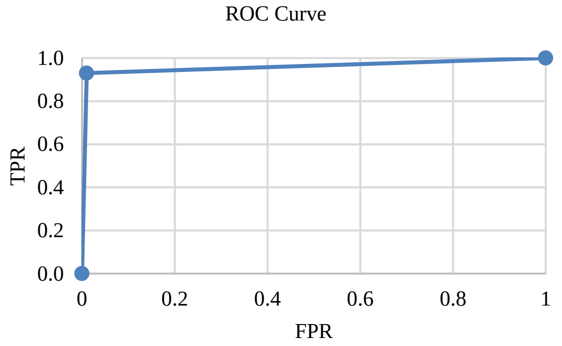

Figure 1. ROC of MPDS model

Figure 2 shows the Precision-Recall (PR) curve depicting the presentation of required MPDS model in conditions of accuracy and evoke is shown in figure. This approach could be experiential because the arc indicates high accuracy and recollection groups. Both precision and recall represent performance levels of the model in identifying churners. High levels in both these metrics indicate that the MPDS model can be used in effective identification of churners.

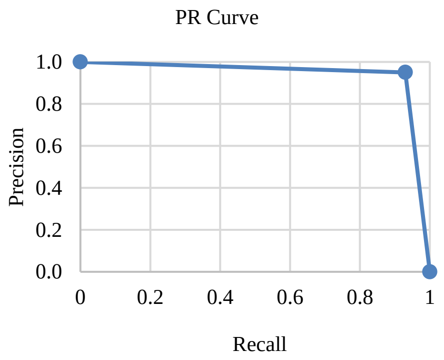

Figure 2. PR Arc of MPDS model

A performance comparison is done with model by Cenggoro, et al., [21]. Comparison is performed based on accuracy and F1-Score. Both are aggregate measures representing the overall performance of the models. Graph in Figure 3 shows that the MPDS model exhibits better performance compared to Cenggoro et al. This indicates the efficiency of performance of the proposed model.

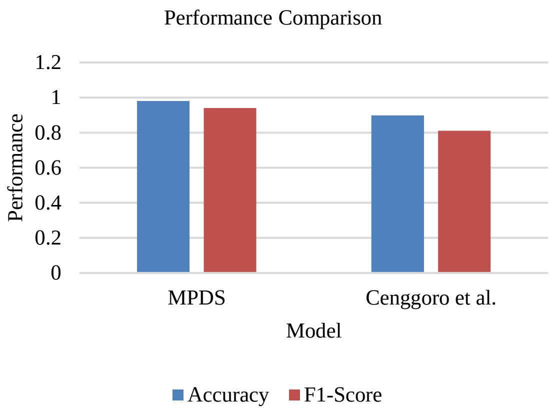

Figure 3. Performance Comparison of MDPS Model

A tabulated view of the performance is presented in Table 4. The MPDS model demonstrations 9% enhanced precision levels, as well as 13% enhanced F_1 connected Score, representing that required MPDS model exhibits a huge advantage over Cenggoro, et al. in effectively identifying churners.

Table 4. Performance comparison of MPDS

	MPDS (proposed)	Cenggoro, et al. [21]
Precision	0.98	0.897
Score and levels of F_1	0.94	0.810

5. CONCLUSION

Churn prediction is a major component in any customer related organization dealing with B2C type of transactions. According to research, gaining a new customer is costlier compared to maintaining an existing customer. Hence, reducing churn levels is highly beneficial to any organization. This work presents the Multi-Level Profit Driven Stacking (MPDS) model for effective prediction of churners. The required model has its significant concentration on enhancing the profit levels of organizations, hence making them usable in real-time. Comparison with existing model in literature indicates that the MPDS model exhibits a huge advantage over the predictions. Prediction accuracy of the MPDS model has been observed to be 11% higher, and F1-Score has been observed to be 13% higher than the compared model. This indicates that the MPDS model exhibits an edge over the model proposed by Cenggoro et al. in effectively classifying the churners. Future enhancements of the model can be based on effective building of training data by identifying and extracting behavior patterns from the base data. This makes the data highly qualitative, hence resulting in better predictions.

REFERENCES

[1] M. Ballings, D. Van den Poel, "Customer Event History for Churn Prediction: How Long is Long Enough? Expert Systems with Applications", Expert Systems with Applications, Vol. 39, Issue 18, pp. 13517-13522, December 2012.
[2] E.H. Xuan Nguyen, "Customer Churn Prediction for the Icelandic Mobile Telephony Market" School of

Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland, September 2011.
[3] J. Hadden, A. Tiwari, R. Roy, D. Ruta, "Computer Assisted Customer Churn Management: State of the Art and Future Trends", Computers & Operations Research, Vol. 34, No. 10, pp. 2902-2917, October 2007.
[4] E. Ascarza, "Retention Futility: Targeting High-Risk Customers Might be Ineffective", Journal of Marketing Research, Vol. 55, Issue 1, pp. 80-98, February 2018.
[5] M. Oskarsdottir, T. Van Calster, B. Baesens, W. Lemahieu, J. Vanthienen, "Time Series for Early Churn Detection: Using Similarity Based Classification for Dynamic Networks", Expert Systems with Applications, Vol. 106, pp. 55-65, September 2018.
[6] W. Buckinx, D. Van den Poel, "Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non-Contractual FMCG Retail Setting", European Journal of Operational Research, Vol. 164, No. 1, pp. 252-268, July 2005.
[7] F. Devriendt, J. Berrevoets, W. Verbeke, "Why you Should Stop Predicting Customer Churn and Start Using Uplift Models", Information Sciences, Vol. 548, pp. 497-515, February 2021.
[8] R. Gubela, S. Lessmann, S. Jaroszewicz, "Response Transformation and Profit Decomposition for Revenue Uplift Modeling", European Journal of Operational Research, Vol. 283, No. 2, pp. 647-661, June 2020.
[9] K. Kane, V. Lo, J. Zheng, "Mining for the Truly Responsive Customers and Prospects Using True-Lift Modeling: Comparison of New and Existing Methods", Journal of Marketing Analytics, Vol. 2, No. 4, pp. 218-238, December 2014.
[10] S. Hoppner, E. Stripling, B. Baesens, S. Broucke, T. Verdonck, "Profit Driven Decision Trees for Churn Prediction", European Journal of Operational Research, Vol. 284, No. 3, pp. 920-933, August 2020.
[11] T. Verbraken, W. Verbeke, B. Baesens, "A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models", IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 5, pp. 961-973, May 2013
[12] N. Alboukaey, A. Joukhadar, N. Ghneim, "Dynamic Behavior Based Churn Prediction in Mobile Telecom", Expert Systems with Applications, Vol. 162, p. 113779, December 2020.
[13] T. Vafeiadis, K. Diamantaras, G. Sarigiannidis, K. Chatzisavvas, 'A Comparison of Machine Learning Techniques for Customer Churn Prediction", Simulation Modelling Practice and Theory, Vol. 55, pp. 1-9, June 2015.
[14] F. Castanedo, G. Valverde, J. Zaratiegu, A. Vazquez, "Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network - PDF Free Download", Docplayer.net, 2014, https://docplayer.net/22913820-Using-deep-learning-to-predict-customer-churn-in-a-mobile-telecommunication-network.html.
[15] S. Hung, D. Yen, H. Wang, "Applying Data Mining to Telecom Churn Management", Expert Systems with Applications, Vol. 31, No. 3, pp. 515-524, October 2006.
[16] M. Ozmen, E. Aydogan, Y. Delice, M. Toksari, "Churn Prediction in Turkey's Telecommunications Sector: A Proposed Multiobjectiv-Cost-Sensitive Ant Colony Optimization", WIREs Data Mining and Knowledge Discovery, Vol. 10, No. 1, October 2019.
[17] Hossein Abbasimehr, Mostafa Setak, M.J. Tarokh, "A Neuro-Fuzzy Classifier for Customer Churn Prediction", International Journal of Computer Applications, Vol. 19, No. 8, pp. 35-41, April 2011.
[18] T. Vafeiadis, K. Diamantaras, G. Sarigiannidis, K. Chatzisavvas, "A Comparison of Machine Learning Techniques for Customer Churn Prediction", Simulation Modelling Practice and Theory, Vol. 55, pp. 1-9, June 2015.
[19] K. Clement, H. Li, C. Wilson, K. Hillary, "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining", IJCSI International Journal of Computer Science, Vol. 10, Issue 2, pp. 1694-1784, March 2013.
[20] A. Omar, F. Hosam, J. Khalid, H. Osama, G. Nazeeh, "Predicting Customer Churn in Telecom Industry Using Multilayer Perceptron Neural Networks: Modeling and Analysis", Life Science journal, Vol. 11, Issue 3, January 2014.
[21] T.W. Cenggoro, R.A. Wirastari, E. Rudianto, M.I. Mohadi, Dinne Ratj, Bens Pardamean, "Deep Learning as a Vector Embedding Model for Customer Churn", Procedia Computer Science, Vol. 179, pp. 624-631, February 2021.
[22] S.M. Kostic, M.I. Simic, M.V. Kostic, "Social Network Analysis and Churn Prediction in Telecommunications Using Graph Theory", Entropy, Vol. 22, No. 7, p. 753, July 2020.
[23] Y. Richter, E. Yom Tov, N. Slonim, "Predicting Customer Churn in Mobile Networks Through Analysis of Social Groups", SIAM International Conference on Data Mining (SDM 2010), Columbus, OH, USA, pp. 732-741, 29 April-1 May 2010.
[24] C.P. Wei, I.T. Chiu, "Turning Telecommunications Call Details to Churn Prediction Data Mining Approach", Expert Systems with Applications, Vol. 23, Issue 2, pp. 103-112, August 2002.
[25] A. Amin, F. Rahim, M. Ramzan, S. Anwar, "A Prudent Based Approach for Customer Churn Prediction", S. Kozielski, D. Mrozek, P. Kasprowski, B.M. Mrozek, D. Kostrzewa (Eds.), "Beyond Databases, Architectures and Structures", BDAS 2015, Communications in Computer and Information Science, Vol. 521, Springer, Cham, 2015.
[26] V. Kavitha, G. Hemanth Kumar, S.V. Mohan Kumar, M. Harish, "Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms", International Journal of Engineering Research and Technology (IJERT) Vol. 9, Issue 5, May 2020.
[27] H. Jain, A. Khunteta, S. Srivastava, "Churn Prediction in Telecommunication using Logistic Regression and Logit Boost", International Conference on Computational Intelligence and Data Science (ICCIDS 2019), Procedia Computer Science, Vol. 167, pp. 101-112, 2020.
[28] "Churn in Telecom's Dataset", 2020 www.kaggle.com/becksddf/churn-in-telecoms-dataset.
[29] K. Sivasaankar, N.P. Gopalan, "Heterogeneous Ensemble Stacking with Minority Upliftment (HESMU) for Churn Prediction on Imbalanced Telecom Data", Materials Today Proceedings, February 2021.

[30] N. Falih, B. Jabir, "Big Data Analytics Opportunities and Challenges for the Smart Enterprise", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 47, Vol. 13, No. 2, pp. 20-26, June 2021.
[31] R. Samsami, "Comparison Between Genetic Algorithm Particle Swarm Optimization and Ant Colony Optimization Techniques for NO_X Emission Forecasting in Iran", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 16, Vol. 5, No. 3, pp. 80-85, September 2013.

BIOGRAPHY

Jude Nirmal Victor was born in Palayamkottai, Tirunelveli, Tamil Nadu, India in 1975. He has a bachelor's degree in computer science and a master's degree in computer science from St. Joseph's College, Bharathidasan University, Trichy, India which awarded him a Ph.D. He is currently a Professor at Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli, India. He has more than 20 years of teaching experiences. He has organized many national and international conferences. He also delivers the guest lecturers about the data analytics and other emerging technologies in various colleges. His current research interests include data analytics, software engineering, big data, data mining and so on.