Notice for the PhD Viva-Voce Examination

Ms Soumi De (Registration Number: 1981905), PhD scholar at the School of Sciences, CHRIST (Deemed to be University), Bangalore will defend her PhD thesis at the public viva-voce examination on Saturday, 14 September 2024 at 10.00 am in Room No. 044, Ground Floor, R & D Block, CHRIST (Deemed to be University), Bengaluru - 560029.

Title of the Thesis	:	**A Model for Churn Prediction Based on Qualitative Support Interaction Features for Hotel Technology Provider**
Discipline	:	**Data Science**
External Examiner (Outside Karnataka)	:	**Dr Prakash P** Associate Professor Vellore Institute of Technology Department of Computer Science and Engineering Chennai, Tamil Nadu
External Examiner (Within Karnataka)	:	**Dr Gururaj H L** Associate Professor Department of Information Technology Manipal Institute of Technology Bengaluru, Karnataka
Supervisor	:	**Dr Prabu P** Associate Professor Department of Computer Science School of Sciences CHRIST (Deemed to be University) Bengaluru 560029 Karnataka

The members of the Research Advisory Committee of the Scholar, the faculty members of the Department and the School, interested experts and research scholars of all the branches of research are cordially invited to attend this open viva-voce examination.

Place: Bengaluru
Date: 02 September 2024

Registrar

ABSTRACT

Customer retention is a significant driver of a company's growth. Machine learning has gained immense popularity as a means to predict customers at risk of churn. Churn prediction models are capable of highlighting customers who are at high risk of churn well in advance. A popular approach to improve the performance of churn prediction models is by using input variables that are mainly quantitative and structured in nature. There are limited works in literature that investigate smart means to effectively utilize and integrate unstructured data into churn prediction models, and study the impact on model efficacy. One of the roadblocks to effectively utilize unstructured data is the associated cost of annotation which is both time consuming and requires intensive manual effort. To overcome this obstacle, researchers often adopt a semi-supervised approach called active learning that aims to achieve state-of-the-art performance using minimal number of samples. Although active learning boosts classifier performance, the underlying query strategies are unable to eliminate redundancy in selected samples for manual annotation. Redundant samples lead to increased cost and sub-optimal performance of learner. Inspired by this challenge, the study proposes a new representation-based query strategy that selects highly informative and representative subsets of samples for manual annotation.

Data comprises messages of a set of customers sent to a service provider. Series of experiments are conducted to analyse the effectiveness of the proposed query strategy, called "Entropy-based Min Max Similarity" (E-MMSIM), in the context of topic classification for churn prediction. The foundation of E-MMSIM is an algorithm that is popularly used to sequence proteins in protein databases. The algorithm is modified and utilized to select the most representative and informative samples. The performance is evaluated using F1-score, AUC and accuracy. It is observed that "E-MMSIM" outperforms popular query strategies, and improves performance of topic classifiers for each of the 4 topics of churn prediction. The trained topic classifiers are used to derive qualitative features. These features are further integrated with structured variables for the same group of customers to predict churn. Experiments provide evidence that inclusion of qualitative features derived using E-MMSIM, enhance the performance of churn classifiers by 5%.

Keywords: churn prediction, active learning, topic classification, uncertainty sampling

Publications:

1. **De, S., & Prabu, P.** (2022). A sampling-based stack framework for imbalanced learning in churn prediction. IEEE Access, 10, 68017-68028.
2. **De, S., & Prabu, P.** (2023). A Representation-Based Query Strategy to Derive Qualitative Features for Improved Churn Prediction. IEEE Access, 11, 1213-1223.
3. **De, S., & Prabu, P.** (2022). Predicting customer churn: A systematic literature review. Journal of Discrete Mathematical Sciences and Cryptography, 25(7), 1965-1985.