Hybrid Learning Model for Anomaly Detection and Automatic Labeling for Heart Disease Prediction.

Nivedhitha.V Assistant professor,
Department of Computer Science and Engineering, SSM Institute of Engineering and Technology, Dindigul, Tamilnadu, India.

Avanthiga.S.K , MahaLakshmi.P ,Vimalakarthika.P

Department of Computer Science and Engineering, SSM Institute of Engineering and Technology, Dindigul, Tamilnadu, India.

ABSTRACT

This project proposes a Hybrid Learning Model which uses both Clustering and Classification methods (HLMCC) to automate the labelling process and detect anomalies in data mining. The model consists of two practical phases, automatic labelling and detecting anomalies. First the HLM groups the data into normal labelled one and unlabelled data clusters by adopting Hierarchical Affinity Propagation (HAP) clustering. Second, the labelled data obtained from the clustering phase is used to train the Decision Trees (DTs) and to classify future unseen data. The results show that the HLM is able to automate the labelling of data, which is beneficial to minimize human involvement.

Keywords:. Hybrid Learning, HAP,DT,HLM.

1. INTRODUCTION

The IoT has been found in several application domains such as smart homes, wearable devices, smart cities, health care, agriculture, transportation, and industrial sectors of industry. IoT devices generate data that may behave inconsistently owing to abnormal or anomaly behavior as a result of attack issues or breakdown in devices, as examples. An anomaly, in this context, means an abnormality in the data that differs from the predicted pattern.

The characteristics of an anomaly are: different from the norm and occurring rarely in the Data Anomaly detection is the technique of identifying rare observations which do not follow the expected behavior. The major technique for performing anomaly detection involves the use of machine learning algorithms. This helps to improve the performance of the system by learning from and using data from previous experiences. There are three types of machine learning task, which are supervised, unsupervised, and semi-supervised learning. Supervised learning trains the model based on predefined labeled data, while unsupervised learning similarities between unlabelled data. Semi-supervised learning deals with partially labeled data to build the model.
Most current anomaly detection systems rely on labeled data which may not be available or it is time-consuming and expensive to produce. In addition, the data collected from IoT devices usually lack the class label and form as unlabelled data. Moreover, the volume of IoT data is growing at an increasingly rapid rate, creating a need to predict, detect, and classify any anomaly for future unseen data. To overcome these limitations, this paper proposes a Hybrid Learning anomaly detection Model that employs Clustering and Classification approaches called HLMCC.

The HLMCC model consists of two functional phases: automatic labeling and detecting anomalies. In the automatic labeling phase, Hierarchical Affinity Propagation (HAP) clustering is applied to