Abstract

In this study, an attempt was made using machine learning techniques to discover knowledge that will assist policy makers in taking decisions that will ensure that the sustainable development goals on Health is met. Agglomerative Hierarchical clustering was used to cluster the states by personnel information (number of doctors, community health workers, nurses and midwives), this was visualized using a dendrogram. The Exploratory analysis revealed that it is only community health workers that are well distributed in all the states, the North West states have the least number of hospitals offering ante-natal services. Random Forest model was used to generate a feature importance to determine the important attributes that determined the availability of maternal health delivery services in a hospital, an important discovery was the fact that the availability of doctors does not in any way determine the availability of maternal health delivery services but rather community health workers, nurses and midwives are the major determinants. Random Forest algorithm was also used to classify hospitals offering maternal health delivery services and the result compared with Logistic Regression, Bagging and Boosting. The evaluation metrics used were accuracy, precision and recall. For accuracy and
precision, Random Forest performed best while for recall it performed poorly compared to all the other algorithms.

References


3. Jiawei Han 2006, Data Mining: Concepts and Techniques, San Francisco, Calif.: Morgan Kaufmann; Oxford: Elsevier Science [distributor], 2nd ed.

4. Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, 1996, From Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence, AI Magazine.


Index Terms

Computer Science  Artificial Intelligence

Keywords

Random Forest, Hospitals, Agglomerative Hierarchical Clustering, Dendrogram