The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2017)

Real-time Miscarriage Prediction with SPARK

Hiba Asri[a,]*, Hajar Mousannif[b], Hassan Al Moatassime[a]

[a]OSER Research Team,FSTG Cadi Ayyad University,Marrakech 40000,Morocco
[b]LISI Laboratory,FSSM Cadi Ayyad University,Marrakech 40000,Morocco

Abstract

Mobile phones and sensors have become very useful to understand and analyze human lifestyle because of the huge amount of data they collect every second. This triggered the idea of taking advantage of such technologies to predict miscarriages and help pregnant women react in advance to a probable miscarriage. To achieve this, we propose to combine benefits and advantages of both machine learning and Big Data tools applied to smartphone/sensors real time generated data. Kmeans clustering algorithm is used for miscarriage prediction and predicted clusters (partitions) are transmitted to the pregnant woman in her front-end user interface in the mobile application, so that she can make quick decisions in case of miscarriage or probable miscarriage. We used real-world data to validate our system and assess its performance and effectiveness. All data management and processing tasks are conducted over Apache Spark Databricks.

Keywords: Miscarriage prediction; Big Data; Data mining; Spark; Databricks; Kmeans.

1. Introduction

Miscarriage is the spontaneous loss of a pregnancy before the 20th week and presents the most common hurtful pregnancy outcome. As miscarriage is an irreversible phenomenon, prediction and prevention are still the only way to deal with this problem, which has negative psychological consequences for the woman and her partner as well[1]. Some studies report that 50 percent of pregnancies will end in miscarriage[2].

Well-established risk factors for miscarriage include history of miscarriage[3], increased maternal age[4,5] and

---

* Corresponding author. Tel.:+212-666-41-79-11; fax: +212-524-43-31-70.
 E-mail address: hiba.asri@ced.uca.ma , asri.hiba@gmail.com

infertility. Also stress and emotional wellbeing remains also important in pregnancy. In fact, the combination between life events and stress degree of the person increase risk of getting miscarriage[6]. Emotions like anxiety, shock, sadness, anger, blame, depression, sleep disturbance have all been described as emotional responses to pregnancy loss[7]. On other aspect, women are asked to participate in regular and moderate exercises while pregnant. Physical activity is defined by using either individual characteristics of a woman's occupation such as standing, stooping, lifting, or walking. However, extreme or strong activity of the body has been associated with an elevated risk of miscarriage[8].

There is also increasing interest in the role that eating well plays in pregnancy[9]. Eating in a restaurant, snack or cafe presents a higher risk of developing certain foodborne illnesses to pregnant women and their unborn children, and that can conduct to a miscarriage. Obesity is also associated with an increased risk of pregnancy complications. Indeed, obesity represents an independent cause of miscarriage[10], with rates that vary between 17 and 27 percent.

The rest of this paper is illustrated over a few sections starting with a brief literature review followed by an overview of miscarriage prediction and detailed prediction process in section 3. Section 4 presents the data conceptual modeling and Spark implementation process; while section 5 presents Kmeans algorithm as data mining tool for clustering data. Section 6 illustrates the environment of the experiment, metrics, required configurations and experimental results. Finally, we conclude the paper by a conclusion and perspectives.

2. Related works

In the past few years, many researches were centered on the use of data mining tools in predicting outcomes. In the specific field of healthcare, data mining shows its power in predicting disease, extract patterns and make decisions. Several healthcare solutions are presented in the markets, which take a value from big data to help people make decisions by themselves to save time[11].

In[12], authors explained the effective and useful sides of big data tools and techniques in healthcare system. They affirm that the integration of big data tools, data mining, medical informatics and big data analytics techniques; is an effective way on healthcare delivery costing and good healthcare outcome. Authors assert that clustering is the most powerful tool used in several forecasting domains[13].

A performance comparison of four main machine learning algorithms: Support Vector Machine (SVM), Decision Tree (C4.5), Naive Bayes (NB) and k Nearest Neighbors (k-NN) on the Wisconsin Breast Cancer (original) datasets is conducted[14]. The main goal of the study is to achieve the correctness in classifying data with respect to efficiency and effectiveness of each algorithm in terms of accuracy and precision. In their research, SVM shows its power in term of performance by reaching a highest accuracy (97.13%) with the lowest error, compared to the use of K-NN, NB and C4.5 classifiers.

In another research[15], underweight, obesity and maternal age are related to increased miscarriage as potentially modifiable pre-pregnant risk factors. The study estimate that miscarriages will decrease by 25.2 percent if we reduce all these risk factors to low risk levels. Also, maternal age and alcohol consumption were identified as the most significant risk factor for miscarriage.

In another research[16], authors present a model for prediction of stillbirth based on medical history and maternal characteristics and evaluate the performance of the model. Logistic regression data mining method is used to determine if there is any contribution to predict stillbirth from several parameters. The most significant characteristics were valued to build a prediction model.

The development of effective predictive and perspective analytics systems relies on the use of technologies such as Big Data, advanced analytics and intelligent systems. In the particular case of miscarriage prediction, studies are usually based on medical history, ultrasound of fetus and/or information from patients. Including real time data