IEICE *Software Interprise Modeling (SWIM)*

Using Biological Variables and Social Determinants to Predict Malaria and Anemia among Children in Senegal

Date: December 2nd, 2017
Venue: Tokyo University of Science
Presented by: M. Boubacar Sow, Miyagi University, Graduate School of Project Design: Information design Program.
Email: p1752007@myu.ac.jp

Outline

- Background and problem Statement
- Purpose
- Malaria and Anemia Definitions
- Disease classification
- Data Collection
- Data Reduction
- Model Building
- Results
- Key Findings
- Research contributions
- Limitations
- Future Work

Introduction

- Machine learning techniques in public health care.
- Improving child health by disease prediction.
- Presenting two classification models for two life-threating diseases.
- Predicting Malaria and Anemia using four machine learning techniques.

Purpose

- The Purpose of this paper is to solve health perspective problems and seeking knowledge for better health.

Hypothesis

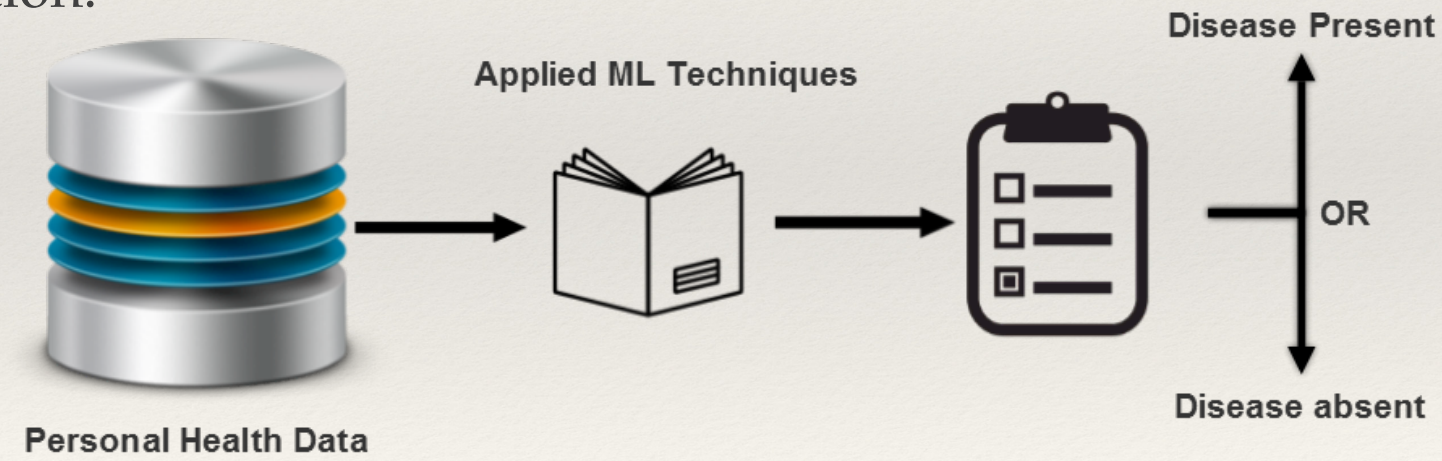
- H1: Using social and biological variables is an innovative approach in solving public health problems by data analytics.
- H2: We can predict Malaria and Anemia by considering data related to mothers as a social determinants of child health
- H3: By data analytics we can figure out the relation between Malaria and Anemia

Malaria and Anemia

- Disorders of the red blood cells:
- Malaria caused by plasmodium parasites from the bite of Anopheles.
- Anemia is caused by iron deficiency.
- Severe Malaria can cause Anemia

What is disease Classification

- The use of Machine learning techniques in Medicine.
- Binary classification: absence or presence of a disease, normal or abnormal condition.



Types of variables used in DC

- Clinical data
- Morphometric parameters
- Physiological data
- environmental data

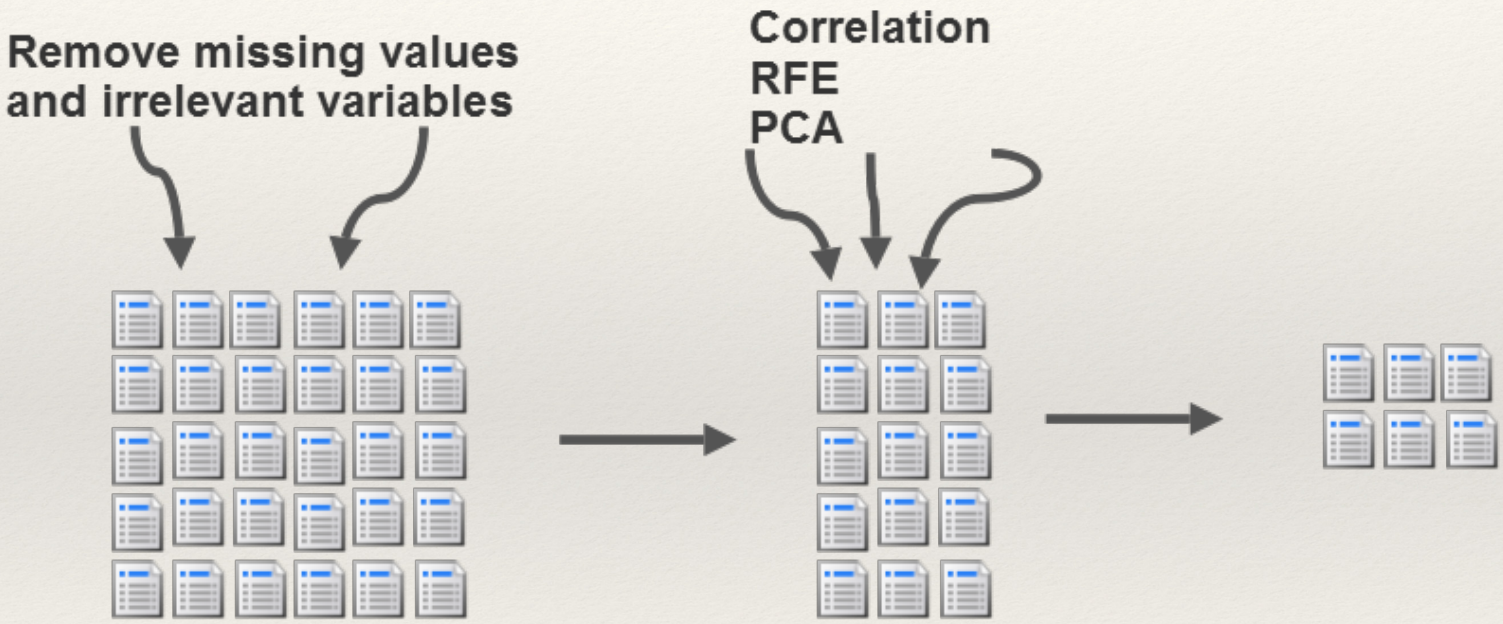
Is it relevant to use SDH and biological variables?

- **Biological variables of health**: Personal biological factors such as Age, Gender, BMI,…
- **Social Determinants of health**: Information related to where people live, what they eat, what they do, what they learn, do they have occupation and those conditions affect their health.

Getting Started by collecting the Data

- Demographic Health survey: Survey made in Senegal 2015 to 2016.
- Initial Dataset : 986 variables and 6935 instances.
- Composition: Household information, Mother's record, Immunization record, Child Information, etc.

Data Reduction



Data Reduction cont'd

Techniques	Number of variables removed	Number of variables remaining
Missing values and irrelevant variables	940	46
Correlation	Anemia(11) Malaria(17)	Anemia(35) Malaria(29)
RFE	Anemia(10) Malaria(5)	Anemia(25) Malaria(24)
PCA	Anemia(5) Malaria(7)	Anemia(20) Malaria(17)

Final Dataset

- Malaria: 62 instances and 17 attributes
- Anemia: 2583 instances and 20 attributes

Binary classification

- supervised learning considering two class labels.
- Separate input data into two classes.
- In medical testing is the approach to determine if a person has certain disease or no.

Binary classification Setting

Anemia

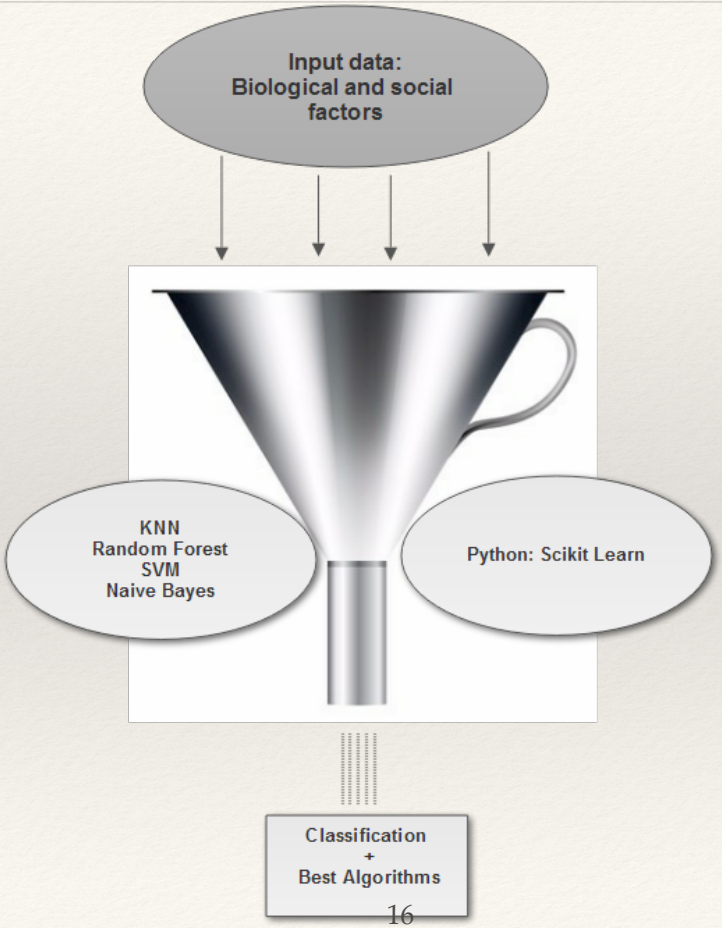
Severe
Moderate → **Positive**
Mild

Not anemic: **Negative**

Malaria

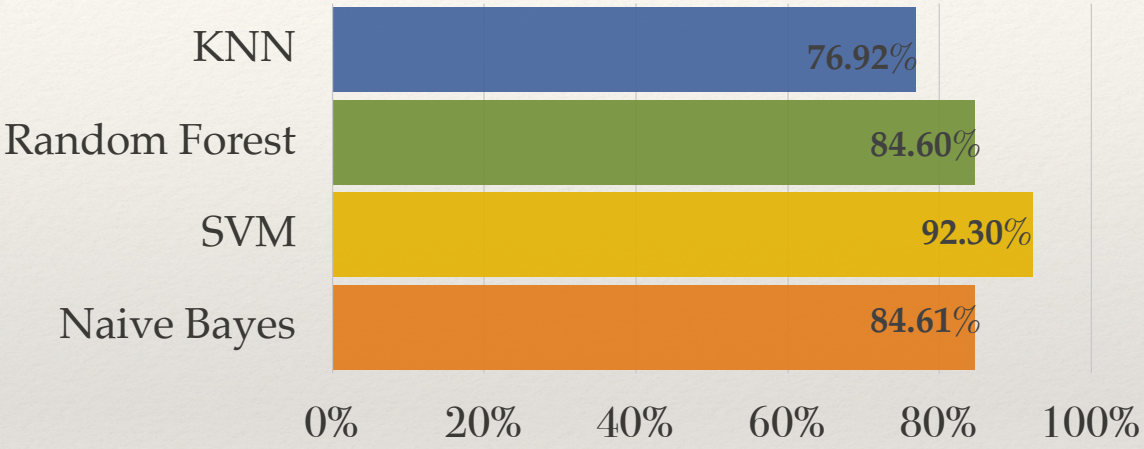
Result of Malaria test:
Positive or **Negative**

Building a Model for Malaria and Anemia



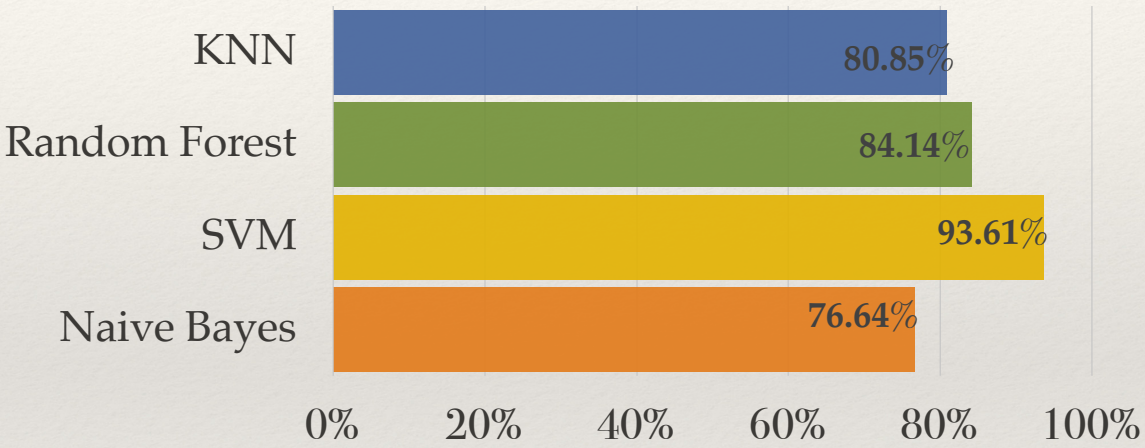
Results

- For Malaria:



Results cont'd

- For Anemia



Key Findings

- SVM Classifies more consistently for two-dimensional output labels(Positive and Negative).
- Non-medical factors predict accurately Anemia and Malaria.
- Information related to mother are good predictors for our classification task.
- Using some common variables for both classification sustain the relationship between Malaria and Anemia.

Research Contributions

- Promote the use of SVM in disease prediction.
- Give more value to Social determinants of health and biological variables for medical diagnostic and healthcare.
- Involving variables related to mother in preliminary disease diagnose.
- This study is a knowledge discovery to tackle child health problems.

Research Limitations

- Parameters tuning for SVM and Random Forests
- This study is limited by the small sample size of Malaria dataset.

Future work

- Classify the severe stage of Malaria known as Malarial anemia.
- Design an Ontology model for Social determinants of Health. (Integration and Interoperability)
- Classification for multi-class problems.
- Use social determinants of heath to predict other diseases.

THANK YOU