**Proiect**

**Aplicatie pentru Centralizare, explorarea si analyza multivariata**

**a datelor rezultate din campaniile de screening col uterin**

**Proiect: Specificatii aplicatie informatica**

**Domeniu: Oncologie**

**Scop: Asistenta si eficientizarea procesului de screening de cancer de col uterin**

**Autor: Andreea ITU**

**Data: 15.03.2017**

**Descriere:**

Prezentul documente constituie prezentarea specificatiilor minime necesare in vederea proiectarii si implementarii unei aplicatii informatice destinate analizei inferentiale a datelor generate in urma procesului desfasurarii campaniilor de screening pentru cancerul de col uterin.

Aplicatia va fi utilizata in atat in vederea inregistrarii datelor generate in timpul campaniei dar mai ales pentru realizarea de analize inferentiale in vederea stabilirii de corelatii intre variabilele observate, realizarea de segmentari ale populatiei testate sau chiar construirea de modele inferentiale avansate (modele care sa permita replanifiicarea campaniilor, luarea de dezicizii curente in viitor, etc).

Principalele elemente referitoare la datele pe care aplicatia le va manipula sunt urmatoarele:

* **Centralizarea** (importare din surse externe) sau introducerea manuala a datelor persoanelor testate precum si a etapelor testarii si rezultatele aferente
* Analiza **demografica** a populatiei testate cu vizualizarea datelor prin plot-uri/grafice bidimensionale cu sau fara harti geografice
* **Analiza** si teste Pearson in vederea stabilirii de **corelatii** intre diversele variabile observate sau variabile derivate – de exemplu analiza corelatiei intre media varstei populatiei testate si diversele zone testate sau corelatia intre etnie, nivelul de dezvoltare al comunitatii locale si rata de incidenta in respectiva comunitate
* **Modele** si **grafice** de **regresie** care sa genereze inferente cu privire la diverse variabile tinta importante in campaniile de screening – de exemplu potentiala prezenta in campania de testare a populatiei pentru o locatia inca nevizitata in functie de informatiile cunoscute si clasificate privitoare la acea locatie tinta

**Etape:**

Principalele etape ale proiectului de construirea a aplicatiei informatice sunt urmatoarele:

1. **Analiza specificatilor** – prevede continuarea si finalizarea prezentului document cu completarea tututor studiilor de caz de utilizare a aplicatiei informatice
2. **Arhitectura aplicatiei informatice** - prevede stabilirea detaliilor in ceea ce priveste mediul in care va functiona aplicatie (de exemplu aplicatie online/web sau aplicatiei instalata in windows), disponibilitatea pe medii mobile, structura datelor, structura interfetelor vizuale (meniuri, ferestre, butoane, functionalitati), precum si algoritmii utilizati pentru diversele functionalitati. Aceasta etapa va contine totodata si propunerile de metode si algoritmi necesari in procesul de analiza inferentiala – cum ar fi cazul predictiei participarii la campania de screening prin utilizarea unui model de regresie multi-variata (in baza experientelor anterioara in variate comunitati se vor putea introduce informatii in modelul de regresie liniara multi-variata care va putea genera o valoarea continua reprezentand predictia participarii intr-o locatie tinta noua si nevizitata – astfel eficientizandu-se planificarea si alocarea resurselor in proiect in vederea maximizarii numarului de vieti omensti salvate)
3. **Dezvoltarea si implementarea aplicatiei** – reprezinta etapa efectiva de scriere a codului sursa atat pentru bazele de date (de exemplu SQL), cat si pentru interfetele grafice (de exemplu C# .NET) precum si pentru modelele de analiza inferentiala si analiza statistica in general (de exemplu limbaj R sau Python)
4. **Testarea si punere in productie** – reprezinta ultima etapa in care aplicatia este lansata efectiv in productie dupa efectuarea unei testari

**Unelte necesare:**

1. **Analiza si Proiectare: UML, Microsoft Visio**
2. **Dezvoltare, implementare si testare: Microsoft Visual Studio, R Studio, Python 3.5**
3. **Biblioteci si module propuse:**
   1. **Interfete grafice: .NET, Ajax**
   2. **Vizualizare grafica: ggplot2, Matplotlib, Bokeh**
   3. **Manipulare date: Dplyr, Pandas**