

## OR7 - Sviluppo di servizi per la Big Data Analytics

### Descrizione sintetica dell'obiettivo

L'Obiettivo Realizzativo di sviluppo sperimentale ha avuto quale risultato principale l'implementazione di algoritmi relativi alla Big Data Analytics per l'erogazione di servizi a valore aggiunto, quali: i) predizioni sul fattore di rischio medico-sanitario dell'utilizzatore in base alla sua storia personale e/o rapportandolo alla conoscenza acquisita dall'analisi dei dati su tutta la popolazione, ii) bollettini statistici "geolocalizzati" dai quali è possibile definire degli indicatori utili per la profilazione della popolazione, iii) sistemi per il monitoraggio di possibili criticità mediche in tempo reale, come gli andamenti e le evoluzioni di epidemie, iv) estrazioni di informazioni funzionali utili alla distribuzione delle risorse medico/sanitarie sul territorio, v) statistiche avanzate per l'elaborazione dei dati, vi) sviluppo di un cruscotto intelligente.

L'OR in esame è stato ottenuto con successo attraverso le attività seguenti.

### A7.1 – Progettazione e sviluppo dei servizi di Big Data Analytics

#### Obiettivo della Attività

In questa attività saranno realizzate tecniche e servizi in modo da garantire scalabilità orizzontale, ossia mantenere livelli di prestazioni costanti al crescere dei flussi di dati da analizzare, attraverso l'utilizzo di file system distribuiti capaci di archiviare ed elaborare grandi volumi di dati indipendentemente dal fatto che essi siano o meno strutturati, seguendo le specifiche precedentemente formalizzate nell'attività A3.1.

Le tecniche di Big Data Analytics sviluppate dovranno supportare in maniera altamente flessibile attività di elaborazione e di analisi adatte per dati clinici e sanitari espressi sia in forma relazionale che non relazionale (ad esempio, dati basati su testo, file di log, dati prodotti da social media, dati prodotti da sensoristica indossabile). I dati dovranno essere aggregati e/o incrociati secondo differenti modalità così da ottenere risposte rapide ed esaurienti per le esigenze dell'utenza.

#### Descrizione Attività

I risultati principali raggiunti nell'attività A7.1 si possono riassumere nell'individuazione di strumenti, tecniche e servizi caratterizzati dal fatto di essere in grado di garantire prestazioni costanti in caso di crescita esponenziale di dati sanitari da elaborare. Più nel dettaglio, le tecniche ottenute hanno permesso di garantire scalabilità orizzontale, ossia di riuscire a mantenere livelli di prestazioni costanti al crescere dei flussi di dati da analizzare, attraverso l'utilizzo di file system distribuiti capaci di archiviare ed elaborare grandi volumi di dati e di strumenti di Big Data Analytics. Inoltre, sono state realizzate tecniche atte alle attività di elaborazione e di analisi richieste specificamente per dati clinici e sanitari fortemente eterogenei, espressi sia in forma relazionale che non relazionale e in diversi formati, quali testi, file di log, dati prodotti da social media, dati prodotti da sensoristica indossabile, immagini e segnali biomedici.

### A7.2 – Implementazione degli strumenti statistici

#### Obiettivo della Attività

L'obiettivo di questa attività sarà la realizzazione degli strumenti statistici definiti nell'attività A3.2. L'attività prevede l'implementazione di un modulo per l'accesso agli insight (predizioni, suggerimenti e reportistica), generati tramite le tecniche di Big Data Analytics precedentemente definite nelle attività dell'OR3. Il front-end sarà realizzato mediante un cruscotto intelligente ed offrirà funzionalità quali: i) statistiche riguardo i dati di wellness; ii) grafici di andamento dei propri parametri vitali e/o di wellness; iii) suggerimenti riguardanti lo stile di vita, basandosi sui dati raccolti.

Saranno inoltre implementate funzionalità di navigazione delle statistiche, in particolare permettendo all'utente di agire su più dimensioni, quali: i) temporali (giorno, mese, anno, etc.); ii) spaziali (comune, provincia, regione, zona, etc.); iii) specifiche della popolazione (sesso, età, provenienza, etnia, etc.).

Per lo sviluppo di questa attività sarà considerato l'uso di tecnologie all'avanguardia come Polymer and AngularJS e di tool specifici per l'analisi statistica, come Cran R, estendendone le funzionalità in base agli specifici requisiti della piattaforma.

#### Descrizione Attività

I risultati raggiunti nell'attività A7.2 vertono nella realizzazione di un'applicazione che permette anche ad un utente meno esperto di riuscire ad interrogare in maniera semplice i dati, per ottenere informazioni e statistiche avanzate relative a dati in ambito sanitario e sullo stile di vita. Nel dettaglio, è stato realizzato un modulo per l'accesso agli insight (predizioni, suggerimenti e reportistica), generati a loro volta sfruttando le tecniche di Big Data Analytics precedentemente definite. Tale modulo è costituito da un front-end che implementa un cruscotto intelligente, realizzato in linguaggio Javascript e con il tool Highcharts. Il cruscotto sviluppato offre la possibilità di ottenere in maniera molto semplice e intuitiva: i) statistiche relative a dati di tipo wellness; ii) grafici sull'andamento dei parametri vitali e/o di wellness; iii) suggerimenti riguardanti lo stile di vita, basati sui dati storici raccolti; iv) navigazione multidimensionale delle statistiche, che permettono all'utente di visualizzare i dati aggregandoli lungo differenti dimensioni, di tipo temporale (giorno, mese, anno, ecc.), spaziale (comune, provincia, regione, zona, ecc.) e specifiche della popolazione (sex, età, provenienza, etnia, ecc.); v) analisi su stili di vita e terapie in corso, per estrarre correlazioni e formulare possibili nessi di causalità atti a formulare suggerimenti automatici riguardanti modifiche allo stile di vita; vi) identificazione di gruppi di persone che mostrano tratti sintomatologici simili, al fine, ad esempio, di creare delle community per mettere in contatto persone con disturbi simili per la condivisione di esperienze terapeutiche e consultazioni mediche.

### A7.3 – Sviluppo dei paradigmi per la realizzazione delle funzionalità di forecasting e inferenziali

#### Obiettivo della Attività

L'obiettivo di questa attività sarà la realizzazione di servizi software comprendenti le tecniche progettate nel corso delle attività precedenti a supporto dell'analisi dei Big Data in particolare per la realizzazione di funzionalità di forecasting e di motori inferenziali.

A tal fine saranno realizzati approcci ibridi che integrano metodi predittivi e di machine learning per l'estrazione di conoscenza da utilizzare nei processi decisionali a partire da dati clinici e sanitari disponibili in forma strutturata e non. In particolare, per applicazioni relative alla classificazione, saranno implementati algoritmi evolutivi, alberi decisionali e/o tecniche statistiche tradizionali ed avanzate specificate nell'attività A3.3.

Diversamente, per applicazioni che necessitano di effettuare una regressione, saranno sviluppate tecniche statistiche specifiche per l'analisi e la modellazione multivariata. Le tecniche che saranno messe a punto consentiranno di trattare situazioni in cui la base di conoscenza risulti non nota a priori oppure definita in maniera vaga e/o incompleta, garantire un alto grado di interpretabilità e, al contempo, massimizzare l'accuratezza della classificazione/regressione ottenuta dal ragionamento clinico modellato.

#### Descrizione Attività

Nell'attività A7.3 sono state completate le funzionalità del cruscotto, frutto delle precedenti attività, realizzando i servizi software dedicati all'estrazione di modelli di classificazione e di modelli previsionali, sfruttando la conoscenza contenuta nella grossa mole di dati a disposizione. Nell'attività, infatti, sono stati implementati i servizi per la realizzazione di funzionalità di forecasting e di motori inferenziali. I moduli che integrano i metodi predittivi sono stati implementati sfruttando metodologie basate su approcci di tipo Machine Learning, Deep Learning e ibridi, con il fine di estrarre dai dati clinici e sanitari disponibili la conoscenza da utilizzare in processi decisionali. Più nel dettaglio, sono stati implementati modelli di classificatori, basati su algoritmi evolutivi, alberi decisionali e tecniche statistiche tradizionali ed avanzate e modelli di regressori, sfruttando tecniche statistiche specifiche per l'analisi e la modellazione multivariata. Le tecniche implementate hanno permesso di trattare situazioni in cui la base di conoscenza non risulta nota a priori oppure definita in maniera vaga e/o incompleta, in modo da garantire un alto grado di interpretabilità e massimizzare l'accuratezza della classificazione/regressione ottenuta dal ragionamento clinico modellato.