Big Data Platform for Public Health Policies

Marco Anisetti, Valerio Bellandi, Marco Cremonini, Ernesto Damiani, Jonatan Maggesi
Dipartimento di Informatica
Università degli Studi di Milano
Crema, Italy, 26013
Email: *firstname.lastname@unimi.it*

***Abstract*—Public Health Policy making process is usually supported by traditional methods to assist the evaluation, like pseudo-evaluation methods and formal evaluation methods, to name but a few. In this paper, we propose the adoption of a Big Data platform supporting the definition and evaluation of a Public Health Policy allowing evidence-based analysis and reducing the need of costly clinical trials. The paper presents a formalized Public Health Policy in deontic logic form, the methodology of the policy making process, and the structure of the Big Data platform implementing the methodology. Concluding, we carried out a preliminary evaluation on a real case study demonstrating the benefits of our approach.**

***Index Terms*—Big Data platform, Public Health Policy**

I. Introduction

Public Health Policies are the fundamental framework of action for a public authority to improve the health of a population. Health policies could be aimed at mitigating or even eradicating epidemics, but more frequently their goal is to act upon an adverse combination of socio-economics and medical factors affecting a disadvantaged subset of the population. For this reason a Public Health Policy often is not just focused on health care, but may include education, psychological, lifesyle, and mobility support. Public Health Policies are expensive, though. Evaluation of costs for the prescribed actions, budget constraints, and cost/benefit analysis are always important factors to consider. On the other side, costs of health care are often a motivation for introducing Public Health Policies with the goal of reducing the incidence of severe diseases or serious health conditions that might require repeated hospitalizations, which are extremely expensive for the public health system.

Given the importance of Public Health Policies for improving the quality of life of many individuals or saving public resources, and the complexity of their definition and evaluation, one relevant research strand is about how technology may support Public Health Policies. We focus, in particular, on Big Data platforms and analytics, studying the way Big Data technology could enhance key aspects of a Public Health Policy process and evaluation.

The definition of Public Health Policies is a complex task requiring an holistic evaluation on a number of heterogeneous data such as medical, behavioral, and environmental data, to name but a few. Currently the process of policy definition mainly relies on the expertise of policy makers and experts, and on the organization of trials to evaluate the effect of a health policy on a population along an extended period of time. That approach suffers of different drawbacks: it is expensive, it is not adaptable, it takes a long time before bringing results, and it does not scale well. Public Health Policies are based on experts knowledge and clinical studies, and once established, collecting evidences to tune them proves difficult. In addition, given an health policy defined at local level (e.g., city or district), evaluating whether it could be extended to a larger scale (e.g. regional or national) is hard to achieve or simulate. The current trend of wearable IoT devices, able to collect health and lifestyle data continuously, and the availability of large dataset of medical data, is creating new opportunities for policy makers to compress costs and counteract the above drawbacks. In this paper, we propose a Big Data based architecture for *i)* helping policy makers in the definition of effective policies, *ii)* continuously evaluating the efficacy of a health policy by monitoring the involved population, and *iii)* supporting the process of transforming a recommendation policy to a prescriptive policy based on the evidence collected about the real efficacy of the policy. In particular, in this work, we sketch a possible integration between a Public Health Policy aimed at reducing the rehospitalization rate for diabetes patients and our Big Data platform providing modern data analysis features. The contribution of the paper is threefold. First, we present a methodology for supporting policy makers in the definition of a Public Health Policy that will exploit the features of a Big Data platform. Then we describe the Big Data platform that we are operating for the definition of Public Health Policy analytics. Finally, a case study based on an open dataset of diabetes patients' clinical data is discussed.

II. Policy Making Process

According to World Health Organization (WHO): "Health policy refers to decisions, plans, and actions that are undertaken to achieve specific health care goals within a society." A health policy is qualified as public if it is made by public institutions for large groups of populations at regional, national or even international level. Loukis et al. [1] define and ontology for a generic public policy workflow and eight stages of a policy definition, each one with specific objectives and a corresponding sub-ontology. This approach is generic but can be adapted to health policy generation based on evidence. These evidences support an analyst in evaluating a policy efficacy. An evidence is traditionally captured through clinical trials or epidemiological studies. According to Dunn et al. [2], a number of traditional methods are also adopted to assist the evaluation like pseudo-evaluation methods, formal evaluation

978-1-5386-0435-9/17/$31.00 ©2017 IEEE

methods, and decision theoretic evaluation.We consider a policy making process as a sequence of subsequent refinements based on three stages: *Situation Analysis*, *Action Plan*, and *Implementation Evaluation and Monitoring*. The approach is derived from [1].

A. Public Health Policy case study

Diabetes, in particular Type-2 diabetes, represents a comorbid medical condition that is associated with rehospitalization [3], [4]. Screening part of the population of patients with diabetes for monitoring existing conditions, such as the glycemic control (HbA1c), and lifestyle, like smoking habit, mobility and level of care, could reduce the readmission rate [5]. A typical Public Health Policy is concerned with the possibility of improving social welfare and the quality of public services; in this case a reduction of readmission rates for patients with diabetes is both an improvement on quality of life for those patients that less often are suffering of diabetic ketoacidosis, and an improvement of hospital management, because costs for a patient with diabetes are estimated as more than double than a patient without diabetes. A Public Health Policy has the problem of evaluating its effectiveness and sustainability, with respect to the goal. The three stage of the policy making process can be instantiated on this example as follows.

- *Situation analysis*: The literature and field experience state that often patients with diabetes are rehospitalized because of poor control of their conditions and absence of changes in lifestyle. The output of this stage is a draft version of the Public Health Policy, which includes the goal of the policy (i.e., to reduce the readmission rate of patients with diabetes) and an initial set of risk factors, typically derived from the literature [6].
- *Action plan*: The first activity typically consists in selecting patients for a trial. Then evaluation criteria should be established, which means to identify a set of relevant risk factors for re-hospitalization and the way to evaluate them. This set could be a refinement of the one identified in Situation analysis based on additional analysis on available data. The output of this stage is a final version of the Public Health Policy ready to be deployed.
- *Implementation, Evaluation and Monitoring*: The policy is deployed as a *recommendation* and after a trial period, depending on the outcome, it could be transformed into a *prescription*, still kept as a recommendation, or even withdrawn. The output of this stage is the final version of the Public Health Policy.

B. Public Health Policy definition

The output of the above policy making process can be condensed in the following example, liberally adapted for presentation purpose from analysis on hospital readmission [5], [7], [8]:

Example 2.1 (Natural language): It is recommended for patients with diabetes having more than *65 years* of age and with annual income less than *15000$* and already hospitalized at least *twice* to receive a post-discharge visit by a diabetes nurse *every month* for *reducing the readmission rate by* $K\%$.

In the example, the age and the income are well-known risk factors for readmission [7]. The income, in this case, is also a criteria for a patient to be admitted to post-discharge assistance program. The condition on the number of hospitalizations discriminates patients below or above a threshold of abnormal number of readmissions. Example 2.1 can be rewritten in a more abstract form by recognizing predicates, which refer to features and have values.

Example 2.2 (Parametric form): It is recommended for people with more than <*age*> and with less than <*income*> and already hospitalized <*number of readmissions*> to receive a post-discharge visit by a diabetes nurse <*nurse visit frequency*> for <*readmission rate reduction*>. Where:

- *normative features*: "age", "income", and "number of readmissions"; *normative values*: "65", "15000$", and "twice";
- *objective feature*: "nurse visit frequency"; *objective value*: "once per month";
- *goal feature*: "readmission rate reduction"; *goal value*: "K%".

With respect to the definition of predicates, goal predicates are typically defined at Situation Analysis stage, for both features and values, because they represent the outcome that makes the policy effective, either in terms of economic sustainability or improved welfare. Objective predicates depends both on constraints defined in the Situation Analysis stage, for example technical or economic constraints, and possibly on analyses carried out during the Action plan stage. Finally, Normative predicates mostly depends on the analyses and correlations carried out during the Action plan.Furthermore, to fully formalize the policy, in this paper, we consider a simplified *Deontic logic* form [9], which let us express concepts like "recommended" or "obligatory".

Definition 2.1 (Deontic logic-based Policy): Let us consider the following policy:

$$\theta = \begin{cases} Policy: & P_1 \rightarrow M(P_2) \\ Goal: & P_3 \end{cases}$$

It defines a policy where P_1, P_2, P_3 are first order logic preposition expressed in terms of normative, objective, and goal predicates, M is a modal operator $\in \{O, R\}$, where O express obligation and R express recommendation [9].

III. Big Data-assisted policy making platform

Big Data technologies, when employed in public policy making, provide a support for human-centric processes, for evaluating large dataset, for managing data streams, and for advanced graphical analyses. Big Data may also help producing fine-grained models, providing a better holistic view suitable for platform like the one of Prasinos et al. [10]. Figure 1 shows an architectural view of our Big Data platform for policy definition and evaluation. It is an abstract view divided into three layers: visualization, data processing, and data acquisition as proposed by Ardagna et al. [11].

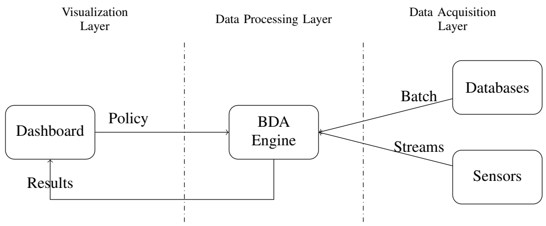

Fig. 1. Big Data platform supporting policy making process.

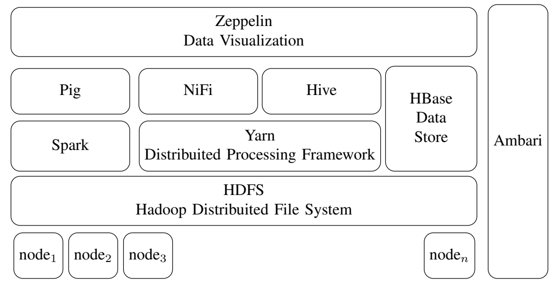

Fig. 2. Big Data platform technologies.

Visualization layer is the frontend for policy makers defining policies and evaluating the results. It is enhanced with features for guided selection and assisted policy specification, in order to foster policy makers autonomy and not requiring the support from Big Data analysts and coding experts.

Data Processing layer is the core of the architecture. It is responsible to process data coming from data acquisition layer. It receives as input the health policy to be evaluated and gives feedback to policy makers on how to tune the policy. The evaluation of a policy is obtained executing a specific analytic pipeline that working on available data provides the result of analysis according to the requirement. Our analytic pipeline specifies all processing activities needed as a set of consecutive steps:

- *data ingestion and preparation* represents the operations to be performed on data before their elaboration. Examples are: anonymization of data, data type conversion or normalization, partitioning;
- *features reduction/selection* describes the features selection from the dataset. For instance, the most significant factors w.r.t. a specific objective are selected;
- *classification* is needed in data mining tasks. Example of classification techniques are clustering, statistical correlation analysis, regression;
- *presentation* is the activity in charge of arranging results for visual analysis.

We note that there are dependencies between each stage of a pipeline, for instance data preparation depends on the classification algorithm adopted (e.g. transformation to numbers), features selection represents a constraint for the preparation.

Data acquisition layer is the layer where data is acquired. Data sources, in general, could be screening or monitoring activities (e.g., through sensors) or clinical databases. These types of data are sensible and need to be treated in accordance with national and international laws and regulation. However, privacy and data protection are out of the scope of this paper.

A. Big Data components

Figure 2 shows the main components of our Big Data platform based on the Apache Foundation ecosystem.

Hadoop - YARN: Hadoop is a tool for data-intensive distributed applications, based on YARN programming model and a distributed file system called Hadoop Distributed Filesystem (HDFS). Hadoop allows writing applications that rapidly process large amounts of data on large clusters of compute nodes. A YARN tool permits to divide the input dataset into independent subsets that are processed in parallel.

HBase: A database engine built on Hadoop and modeled after Google's Big Table. HBase is optimized for real-time data access to large tables in the billions of rows. Among other features, it offers support for interfacing Hive.

Hive: A data warehousing infrastructure, which runs on top of Hadoop. Hive provides a language called Hive QL to organize, aggregate, and run queries on dataset. Hive QL is similar to SQL, it uses a declarative programming model and results are described in one big query. HQL queries can be broken down by Hive to communicate to MapReduce jobs executed across a Hadoop cluster.

Spark: A general purpose cluster computing engine providing APIs to various programming languages such as Java, Python, or Scala. Spark is specialized at making data analysis faster, it supports in-memory computing that enables it to query data much faster compared to disk-based engines such as Hadoop, and also it offers a general execution model that can optimize arbitrary operator graph. Spark also offer several tools, such as machine learning tool M-Lib, structured data processing, Spark SQL, graph processing tool Graph X, stream processing engine called Spark Streaming, and Shark for fast interactive question device.

Zeppelin: A web based and multipurpose notebook that enables interactive data analytics. The notebook is the place for data ingestion, discovery, analytics, visualization and collaboration. Zeppelin supports many interpreters such as Apache Spark, Python, JDBC, Markdown and Shell.

IV. METHODOLOGY

Our methodology implements the Public Health Policy making process described in Section II with the support of the Big Data architecture showed in Section III. We make some assumptions. The first is that relevant data needed for reaching a decision is available on the Big Data platform. The second is that policy makers' decisions are based on historical data present in the platform (*retrospective data*), and that this dataset is comparable to those adopted in longitudinal studies. Data collected in the Big Data platform after the Public Health Policy has been released are considered for policy evaluation and monitoring (*prospective data*).

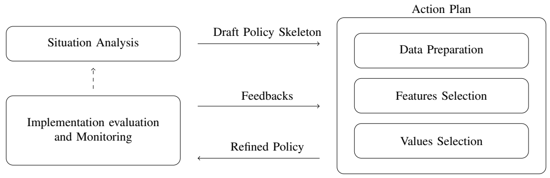

Fig. 3. Policy making process methodology.

Figure 3 shows our methodology in terms of macro stages. In the following, we describe our methodology using the running example introduced in Section II-B as a reference for possible policies.

A. Situation analysis

This stage is mostly concerned with policy makers and domain experts deciding the main features and constraints of a Public Health Policy. Policy makers decide which type of intervention could be needed (e.g., periodic screening, home medical assistance, education, lifestyle monitoring) and the objective of the intervention (e.g., frequency of periodical visits, medication, or medical tests). The output at this stage is a draft version of the policy in a form that will be further refined in subsequent stages. Using our Big Data platform, policy makers and clinicians can inspect retrospective data analyzing features (e.g., behavioral or clinical data, medical exams) to help them drafting the initial Public Health Policy. The role of Big Data engine here is to provide query functionality for an initial policy definition, mostly based on policy makers' experience and available scientific literature.

Example 4.1 (Draft Public Health Policy): Considering Example 2.2 and the formal Definition 2.1, the draft Public Health Policy, with patient p belonging to the population P of hospitalized diabetes patients, can be defined as follows:

$$\vartheta = \begin{cases} Policy: & \forall\ p \in P:\ a_1 \odot\ a_2 \odot \cdots a_n \\ & \rightarrow R(assistanceFreq(p, x)) \\ Goal: & readmissionReduct(P, K) \end{cases}$$

with:

- $\odot$ as a logic operator $\in \{\wedge, \vee, \neg\}$;
- $R(assistanceFreq(p, x))$ the modal operator expressing the recommendation applied the predicate representing the diabetes nurse visiting patient p with frequency x;
- $readmissionReduct(P, K)$ the predicate representing the reduction of $K\%$ in readmission rate of population P.

In the draft version of the Public Health Policy, we see the goal predicate $readmissionReduct(P, K)$ specified with respect to the reference population P and a certain (minimum) outcome $K\%$, reflecting the general goal of improving social welfare and hospital management. The objective predicate $assistanceFreq(p, x)$ is defined with respect to each patient p selected for obtaining assistance and with value x representing the intensity, possibly variable, of the support offered by the health policy. Normative predicates are not necessarily specified at this stage and will be subject to Big Data-oriented analysis.

B. Action plan

This stage is mostly focused on tuning the draft policy evaluating its efficacy using retrospective data. This is an iterative phase with policy makers asking to Big Data platform multiple execution of analytics until a satisfy result is obtained. More specifically, given the draft policy, data preparation are first applied to retrospective data (see Figure 3), then features are selected, and finally features' values are defined. The Big Data platform, at this stage, supports policy makers in defining a suitable preparation of retrospective data and inspecting the results. For instance, data could be groped for comorbidities, personal traits like ethnicity (a well-known risk factor for diabetes), or rehospitalization history (i.e., data partitioning). Statistical analyses can also be performed to evaluate the representativeness of data partitioning (e.g., PCA analysis). When a satisfying preparation of retrospective data for the evaluation is defined, our methodology proceeds with the selection of normative features over retrospective data (feature selection in Figure 3). A set of relevant n features, with respect to the policy goal, should be selected on the prepared retrospective data by means of an analytic approach (e.g., Rough Set). Iteratively, policy makers evaluate the n features and decide whether to further modify them, change the analytic algorithm, or accept the setting. Given the selection of normative features, the methodology proceeds to the definition of normative values (i.e., value selection in Figure 3). In this case, a suitable classification approach should be selected (e.g., regression, clustering, random forest, etc.). With normative features and values, retrospective data can be processed as a training set. The result is a classification that policy makers should evaluate with respect to the policy goal. The decision could be to adopt a different classification approach, modify the normative value selection, or confirm both values and classification approach. Normative features and values could be further evaluated by policy makers, after the definition of normative predicates. If needed the policy maker can restart the process from the data preparation phase. At the end of this stage, the refined Public Health Policy is available for the implementation evaluation and monitoring stage.

Example 4.2 (Refined Public Health Policy): Considering Example 2.2 and Example 4.1, one possible refined policy θ is as follows.

$$\theta = \begin{cases} Policy: & \forall\ p \in P:\ ageGT(p, 65) \wedge \\ & incomeLT(p, 15000) \wedge hospitalGE(p, 2) \\ & \rightarrow R(assistanceFreq(p, month)) \\ Goal: & readmissionReduct(P, K) \end{cases}$$

with predicate $ageGT$ meaning the age greater then a value, $incomeLT$ for the income less then an amount, and $hospitalGE$ for the number of hospitalizations greater or equal to a value.

C. Implementation evaluation and monitoring

During this phase the policy is evaluated on the perspective data as soon as the data is gathered into the Big Data platform. The Big Data platform permits to continuously execute the classification and the comparison with normative and objective features in order to evaluate the efficacy of the policy. When enough positive evidence is obtained, if needed, the modality of the policy can be moved from "recommendation" to "prescription".

From the perspective of policy makers, the support provided by the Big Data approach let them collect real-time evidences of the policy outcome and trends in the observed population of patients. This could be extremely valuable in situations like the rehospitalization of diabetes patients, whose incidence depends on a wide range of factors, many of them correlated with patient's lifestyle, alimentary habits, cultural background, or environmental factors. Providing individual post-discharge assistance and education could sensibly reduce the rate of readmission for patients that lack the willingness for autonomously adopting healthy practices, but the resources (nurses, diabetes specialists) are limited and should be employed for individuals whose readmission risk is very high.

V. Preliminary Evaluation

In this section, we present a walk through evaluation of a simplified policy making process taken from the case study presented in Section II. The aim is to describe our overall approach and methodology with the help of some data. The dataset is a publicly available database [12] with clinical data of 72k US citizens affected by diabetes, which have been hospitalized during a 10 years period. To describe our approach, we suppose that a Public Health Policy, as the one previously discussed aimed at reducing the rate of readmission, has been defined. We also assume that the dataset represents the population of diabetes patients that, starting from a certain date, has been selected as the target of the Public Health Policy. In the following, we describe an execution of the policy making process as supported by our platform. The purpose of the example is to demonstrate how analytic computation can help in the process of informing decisions of policy makers. Let us start by considering the policy drafted in Example 4.1 and a policy maker connected to the Big Data platform to retrieve information for specifying policy predicates. Let us assume that frequency x of predicate $assistanceFreq$ is given (e.g. one visit per month), then the objective of the policy maker is to select a suitable set of normative features and values (Action Plan). The policy goal is to reduce the readmission rate of diabetes patients by $K\%$ ($readmissionReduct$) on a certain time scale (e.g. annually). The evaluation is just preliminary, the goal features are intentionally not taken into account in the following evaluation. Figure 4 shows an algorithmic view of the Big Data pipeline implementing our methodology for the Action Plan stage of this scenario.[1] First, retrospective data is prepared for the evaluation (function $preparation$ in Figure 4), with respect to possible inconsistencies among data types and classification labels. The target feature of readmission had three levels of < 30, > 30 (patient readmitted within 30 days or after 30 days), and no if patient was not readmitted. We re-codified the feature has binary value: patient is readmitted and patient is not readmitted. We obtained 54745 one-time patients and 16773 readmitted patients. Next, we partitioned this dataset considering all the data as retrospective data. Using retrospective data, we started the core of the Big Data processing for feature selection (function $featureSel$ in Figure 4). With respect to labels defined in data preparation and available features in the dataset, advanced feature selection approaches like Rough Set Analysis or Least Angle Regression (LARS) could be used. For simplicity, in this example we selected two features based on medical literature and on the work in Munnangi et al. [13]. Selected features are "max_glu_serum" and "A1Cresult". With these features, the column of the dataset could be reduced by an horizontal cut (function $horizontalCut$). We note that Munnangi et al. [13] uses a number of features to increase the quality of the detection but in this evaluation we are not focused on improving the quality of the policy efficacy but we just want to show the applicability and usability.

After the feature selection, a policy maker should evaluate the quality of the selection in terms of classification provided by those features respectively for one-time and readmitted patients. This classification (Function $classify$ in Figure 4) is obtained using retrospective data as the training set and training a Support Vector Machine (SVM) for patient classification. We assumed that the hyperplane defined by the SVM represents a suitable discriminant for normative values. Given this classification a policy maker can use our Big Data platform to plot the quality obtained. In this case it is quite straight forward that the classification is not satisfactory since it shows an area under RoC curve of 0.54%. For simplicity let us continue with these two normative features.

The policy requires values for each features, therefore to find suitable values for every features (dimensions of the hyperplane), we consider an iterative approach. For each feature, a subset of all value combinations is evaluated with respect to the RoC curve produced if the specific values were used for the classification as they are correlated in the policy. This selection of values (function $featuresVal$ in Figure 4) is then employed to find the best suitable combination of values for normative features (function $projHp$) considering the RoC curve obtained using them instead of the SVM model. The evaluation leads to an initial policy where the normative preposition is as follows: "max_glu_serum' $\neq$"no" $\wedge$ "A1Cresult" $\neq$"no". These values provide precision equal to 37% and recall equal to 51% which confirm that the features are not alone suitable for profitable classification. The rest of the process would verify if using the perspective data, these normative features are suitable to classify patients so that readmission will decreases thanks to home assistance.

As a final remark, a Big Data platform helps the policy

[1] Portions of the implemented script runnable on our architecture is available at https://goo.gl/i5r8Xg

INPUT
$Data$: retrospective labeled patient's data
F_o:Set of objective features
n: number of normative features

MAIN
/*Data preparation*/
P = **preparation**(Data,F_o);
/*Suggested features*/
$F_s = \emptyset$;
/*Manual settings*/
$retry = False$;
repeat
/* – Begin Big Data pipeline – */
[P_f, F] = **featureSel**(P, n, F_s);
$model$ = **classify**(P_f);
[F_v] = **featuresVal**($model$, F);
[$retry$, F_s]= **manualCheck**(P_f,
...F, $model$, F_v);
/* – End Big Data pipeline – */
until ($retry$)
$policy$ = [F,F_v];

PREPARATION(*Data*,F_o)
/*Preparation query*/
P = **DataTypeSetting**($Data$);
P = **Conversion**(P,F_o);
P = **verticalCut**(P,1);
return P

OUTPUT
$policy$: refined policy

FEATURESEL(P,n,F)
/*n number of normative features*/
F_r=**roughSet**(P,n);
if $\neg isempty(F)$
F=$F_r \cup F$
P_f=**horizontalCut**(P, F);
return [F, P_f];

CLASSIFY(P_f)
/*SVM classification */
$model$=**SVM**(P_f);
return $model$;

FEATUREVAL(*model*, F)
/*find values for each*/
/*normative features*/
for each $f \in F$
$out.add$(**RoC**(**projHp**($model$,f));
return selectOptimal(out);

MANCHECK(P_f,.F, $model$, F_v)
/*Visualization and feedback*/
VisualizeFeatures(P_f,.F);
CompareROC($model$, F_v);
AskTuning();
return [$retry$, F_s];

Fig. 4. Action plan Big Data pipeline in algorithmic form.

evaluation through fast evaluations compared with traditional approaches and with effective visualization tool (Function $manCheck$ in Figure 4).

VI. Related Work

Public health policies, in the form of laws, regulations, and guidelines, have a profound effect on public health. However, there is a considerable gap between what research shows as effective and the policies that are enacted and enforced. Research is most likely to influence policy development through an extended process of communication and interaction [14]. A number of methods and techniques can assist analysts in evaluating health policy like pseudo-evaluation, formal evaluation, and decision theoretic evaluation [2]. Their usage requires data sciences knowledge and expertise in the field of application, in addition their integration into a usable policy making-specific framework is far from being realized due to interdependency and heterogeneity of both techniques and programming languages. Big Data analytics enables the capture of insights from data gathered from research, clinical care settings, and operational settings to build evidence for improved care delivery. As indicated in the Institute of Medicine (IoM) report, there are some open problems [15] (e.g. how to manage the data coming from the IoT devices). There is a significant opportunity to improve the efficiencies in the healthcare industry by using an evidence-based learning model, which can in turn be powered by Big Data analytics [15]. Overall, McKinsey & Company estimates that $300 billion to $450 billion can be saved in the healthcare industry from Big Data Analytics [16]. As the focus shifts from cure to preventive health and as new technologies such as wearable sensors evolve as part of the Internet of Things (IoT), the volume of data in healthcare is expected to grow significantly and can provide a wealth of actionable information. The combined power of information from real-time devices, people, clinical systems, and historical population data makes Big Data a very helpful tool in improving the healthcare system [16].

VII. Conclusions

BigData architecture is suitable for policy making because, on contrary to most of the data mining approach, it allows, i) timely response allowing interactions with policy makers, ii) execution of mining task in parallel, iii)supporting automatically monitoring and evaluation of a given policy. In addition BigData allows to take into consideration a number of data and great variety making the classification models more precise.

Acknowledgments

This work was partly supported by EU-funded project EVOTION (contract n. H2020-727521) and by the program "piano sostegno alla ricerca 2015-17" funded by Università degli Studi di Milano.

References

[1] E. N. Loukis, "An ontology for g2g collaboration in public policy making, implementation and evaluation," *Artificial Intelligence and Law*, vol. 15, no. 1, pp. 19–48, 2007.

[2] W. N. Dunn, *Public policy analysis*. Routledge, 2015.

[3] A. D. Association *et al.*, "Screening for type 2 diabetes," *Diabetes care*, vol. 27, no. suppl 1, pp. s11–s14, 2004.

[4] N. Calonge, D. B. Petitti, T. G. DeWitt, A. J. Dietrich, L. Gordis, K. D. Gregory, R. P. Harris, G. Isham, R. Leipzig, M. L. LeFevre *et al.*, "Screening for type 2 diabetes mellitus in adults," *Annals of Internal Medicine*, vol. 148, no. 11, pp. 846–854, 2008.

[5] K. M. Dungan, "The effect of diabetes on hospital readmissions," 2012.

[6] S. M. Haffner, "Epidemiology of type 2 diabetes: risk factors," *Diabetes care*, vol. 21, no. Supplement 3, pp. C3–C6, 1998.

[7] D. J. Rubin, "Hospital readmission of patients with diabetes," *Current diabetes reports*, vol. 15, no. 4, pp. 1–9, 2015.

[8] K. Dungan, S. Lyons, K. Manu, M. Kulkarni, K. Ebrahim, C. Grantier, C. Harris, D. Black, and D. Schuster, "An individualized inpatient diabetes education and hospital transition program for poorly controlled hospitalized patients with diabetes," *Endocrine Practice*, 2014.

[9] P. Balbiani, "Logical approaches to deontic reasoning: From basic questions to dynamic solutions," *International Journal of Intelligent Systems*, vol. 23, no. 10, pp. 1021–1045, 2008.

[10] M. Prasinos, G. Spanoudakis, and D. Koutsouris, "Towards a model-driven platform for evidence based public health policy making," in *Proc. of SEKE2017*, 2017.

[11] C. A. Ardagna, V. Bellandi, M. Bezzi, P. Ceravolo, E. Damiani, and C. Hebert, "A model-driven methodology for big data analytics-as-a-service," in *Proc. of IEEE Congress on Big Data*, 2017.

[12] B. Strack, J. P. DeShazo, C. Gennings, J. L. Olmo, S. Ventura, K. J. Cios, and J. N. Clore, "Impact of hba1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records," *BioMed research international*, 2014.

[13] G. C. Hephzibah Munnangi, "Predicting readmission of diabetic patients using the high performance support vector machine algorithm of sas enterprise miner," in *Proc. of SAS Global Forum*, April 2015.

[14] R. C. Brownson, J. F. Chriqui, and K. A. Stamatakis, "Understanding evidence-based public health policy," *American Journal of Public Health*, vol. 99, no. 9, pp. 1576–1583, 2009.

[15] S. Rickwood, M. Kleinrock, M. Nunez-Gaviria, S. Sakhrani, and M. Aitken, "The global use of medicines: outlook through 2017," *IMS Institute for Healthcare Informatics*, vol. 5, 2013.

[16] D. K. Peter Groves, Basel Kayyali and S. V. Kuiken, *The big data revolution in healthcare, Accelereting value and innovation*. Center fo US Health System Reform, 2013.