Big Data Analytics in Healthcare

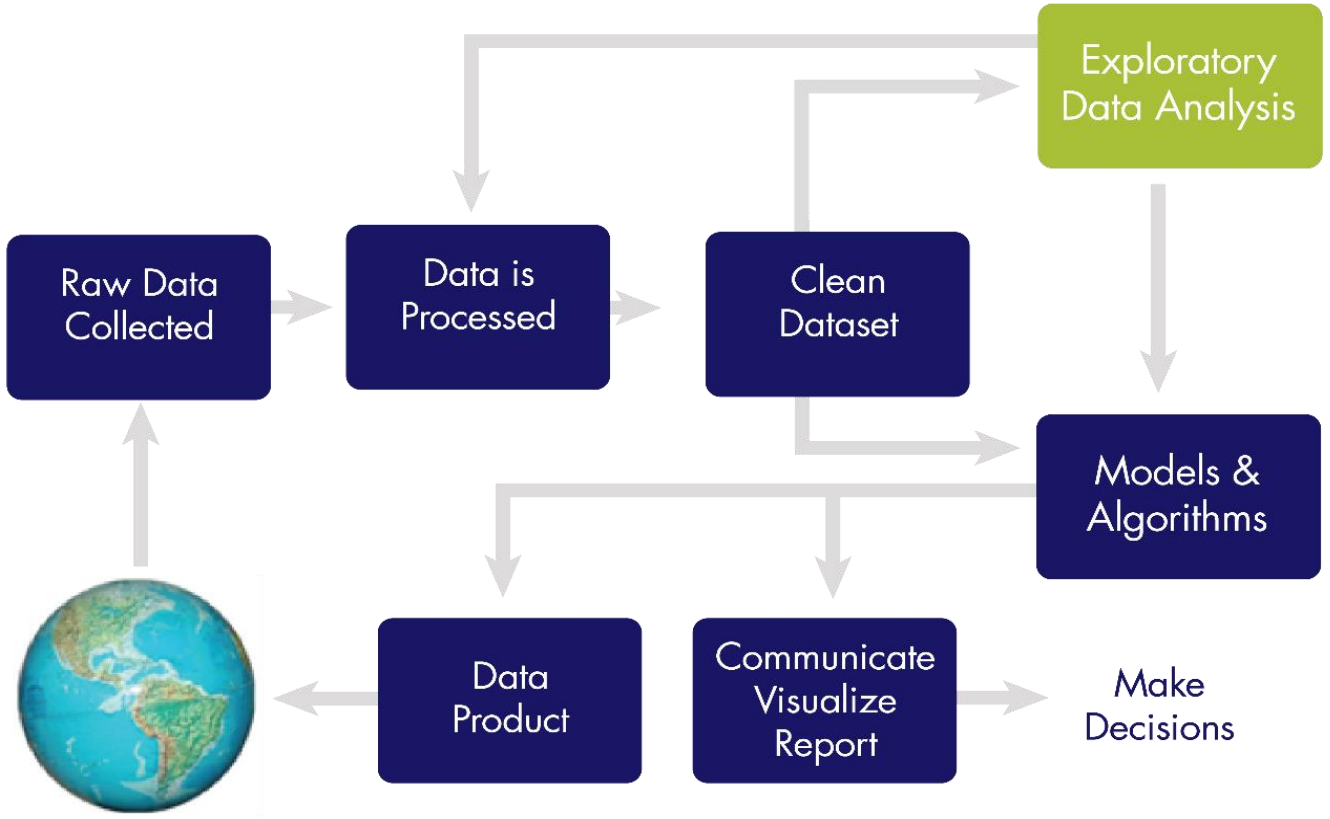
Tsolakidis Anastasios, PhD
Business Intelligence Analyst at
e-Government Center for Social Security(IDIKA SA)

What is analytics?

"The **discovery** of meaningful patterns in data is one of the steps in the data life cycle of collection of raw data, preparation of information, analysis of patterns to synthesize knowledge, and action to produce value."

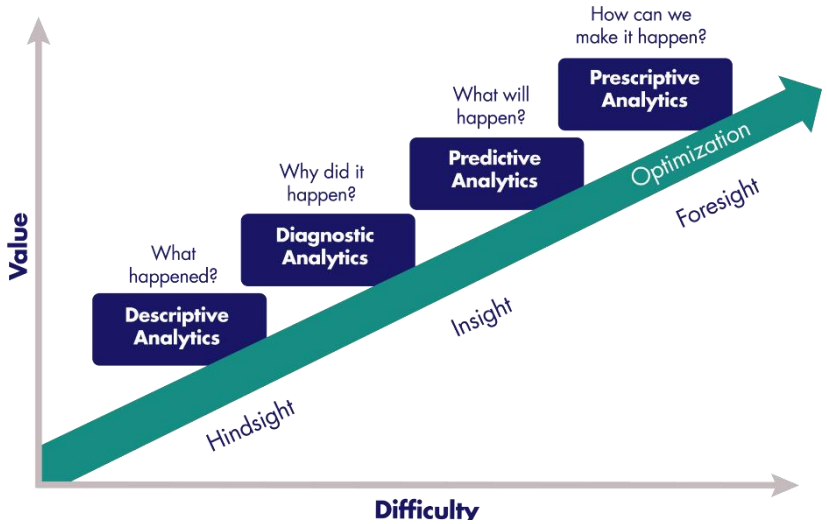
(National Institute of Standards and Technology [NIST], 2015)

The process

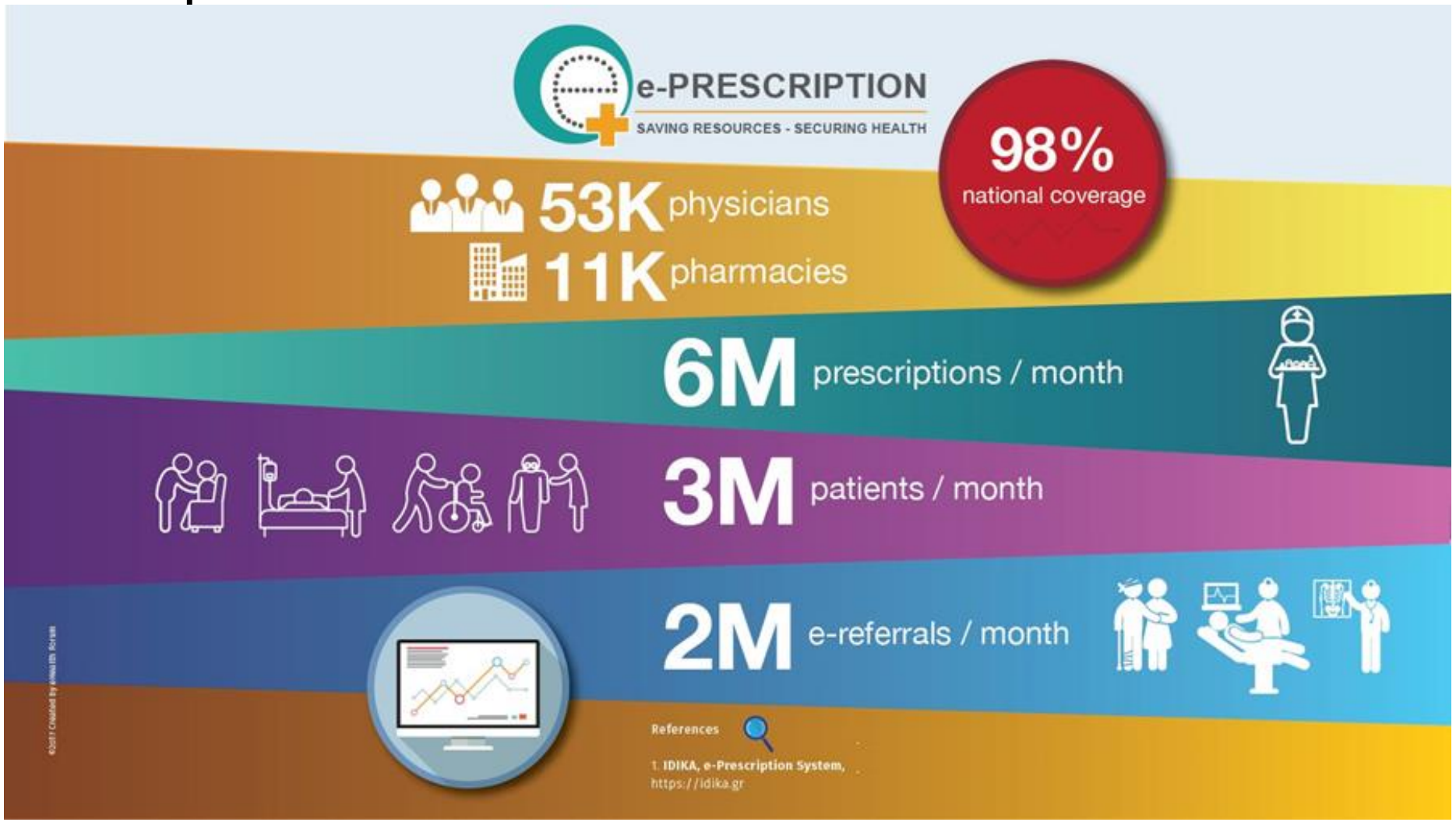

Exploratory Data Analysis
Raw Data Collected
Data is Processed
Clean Dataset
Models & Algorithms
Data Product
Communicate Visualize Report
Make Decisions

Different Types

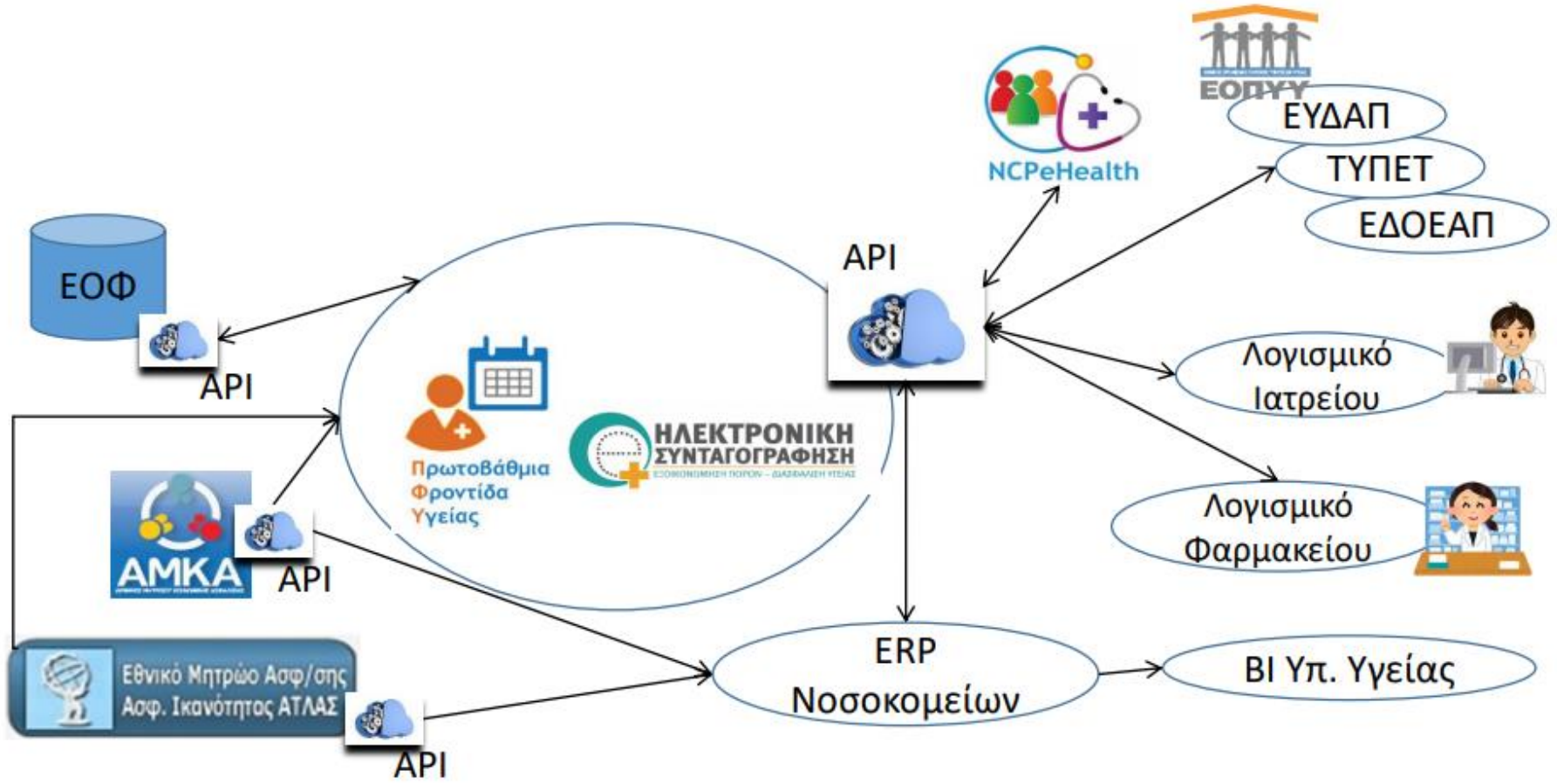
- ***Descriptive:*** Uses business intelligence and data mining to ask: “What has happened?”
- ***Predictive:*** Uses statistical models and forecasts to ask: “What could happen?”
- ***Prescriptive:*** Uses optimization and simulation to ask: “What should we do?”
- ***Diagnostic:*** Examines data to answer “Why did it happen?”



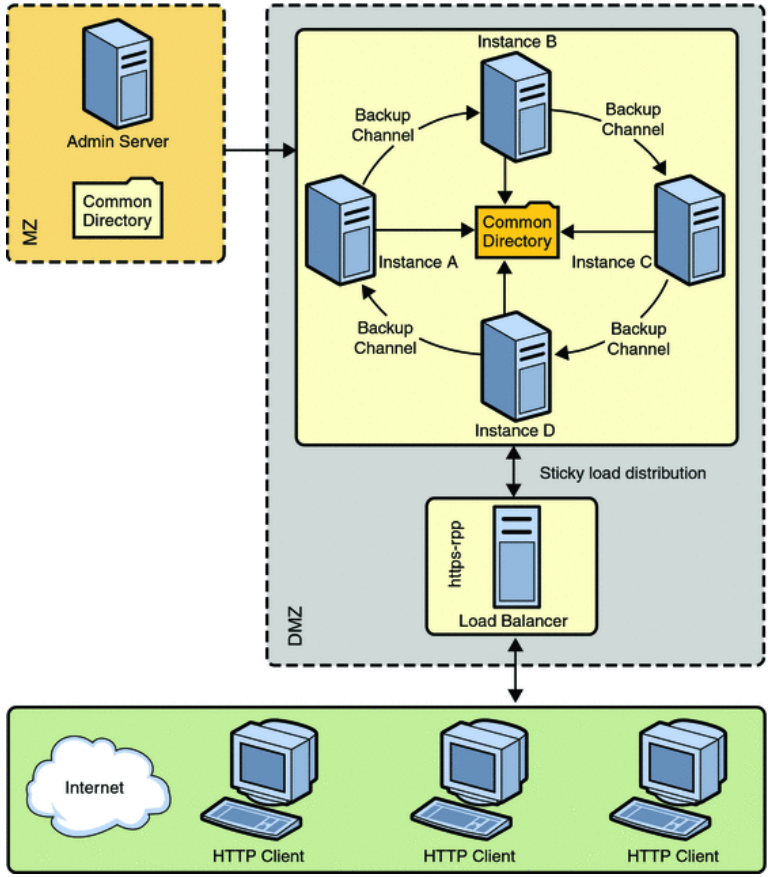
e-Prescription



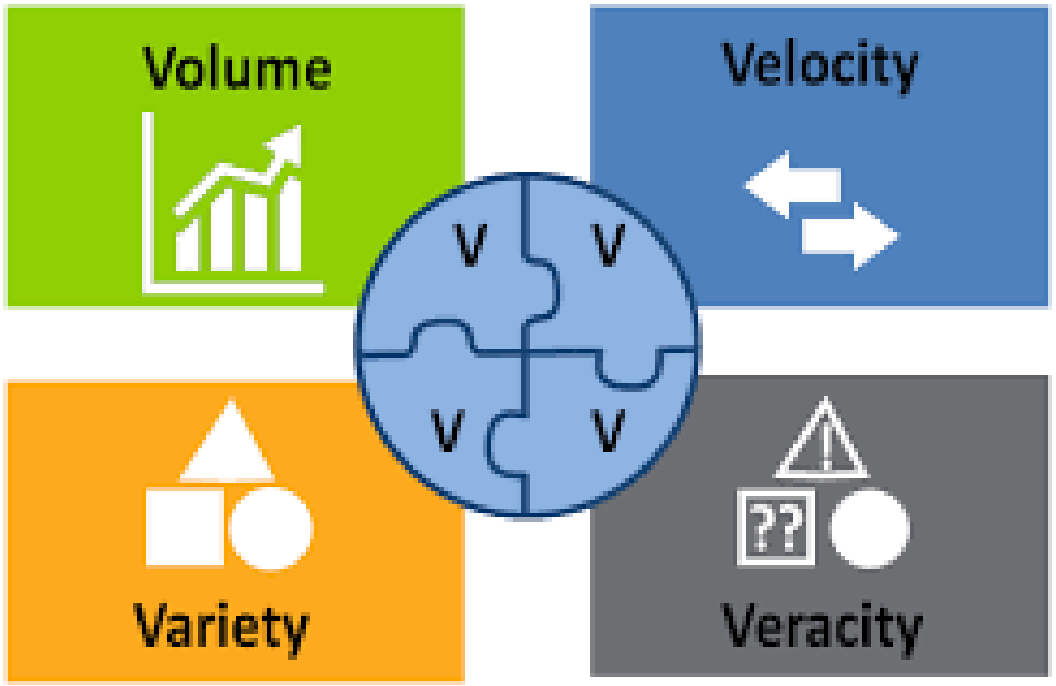
e-Prescription



Database Topology



4 V's of Big Data



Volume

HOW MUCH DATA IS THERE?

- 720.000.000 Prescriptions
- 300.000.000 Exam Prescriptions
- 12.000.000 Patients
- 50.000 Doctors
- 20.000 Pharmacies

Variety

HOW MANY DIFFERENT TYPES OF SOURCES ARE THERE?

- We keep only Structured data

- The following Standards are used
  - ICD10 (International Classification of Diseases (ICD)),
  - ATC (Anatomical Therapeutic Chemical (**ATC**) Classification System)

Velocity

HOW QUICKLY THE DATA CAN BE PROCESSED?

- Some activities are very important and need immediate responses.
- For time-sensitive processes such as fraud detection, Big Data flows must be implemented and scheduling process should be applied.

Veracity

CAN WE TRUST THE DATA?

- Improper use of Diagnosis
  - Noisy Data
  - Outlier Detection

Who Benefits From Big Data Analytics?

- Providers (Clinics, Hospitals). The insights generated from big data analytics enables healthcare providers, such as clinics and hospitals, to improve patient care.

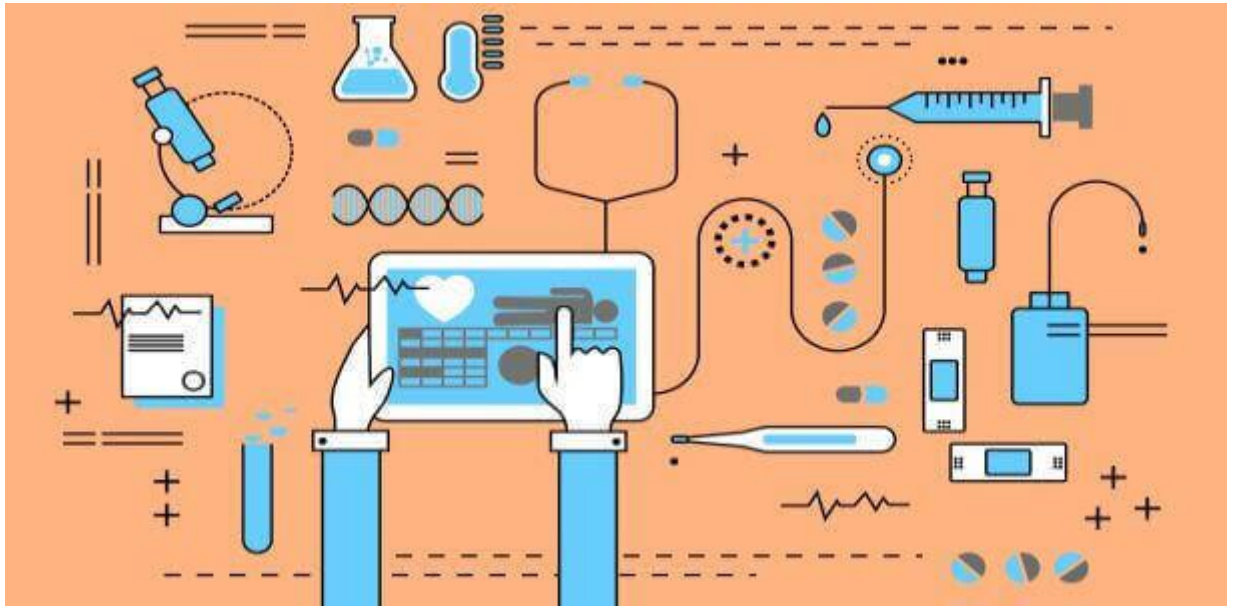
- Payers (Insurance). They can use it to reduce fraudulent activity, rectify false claims, provide better service to their customers and reconcile records faster.

Who Benefits From Big Data Analytics?

- Patients. Patients get better overall care, live healthier lives, save money on insurance and so much more.

- Pharma. More effective drugs and shorter production times. Pharma companies will also save on the costs related to drug development because the process for determining which drugs are worthwhile to enter clinical trials will be more accurate.

Examples of Big Data Analytics In Healthcare



Using Health Data For Informed Strategic Planning

- Care mangers can analyze check-up results among people in different demographic groups and identify what factors discourage people from taking up treatment.
- University of Florida made use of Google Maps and free public health data to prepare heat maps targeted at multiple issues, such as population growth and chronic diseases. Subsequently, academics compared this data with the availability of medical services in most heated areas. The insights gleaned from this allowed them to review their delivery strategy and add more care units to most problematic areas.

Prevent Medicine Abuse

- Using Historical data we can predict, the demand on medicine

Predictive Analytics In Healthcare

- The goal of predictive analytics is to help doctors make data-driven decisions and improve patients' treatment.
- This is particularly useful in case of patients with complex medical histories, suffering from multiple conditions.

Predict the number of Staff Based on the amount of patients

- Hospitals can use data from a variety of sources to come up with daily and hourly predictions of how many patients are expected to be at each hospital using “time series analysis” techniques.
- These analyses allowed the researchers to see relevant patterns in admission rates. Then, they could use machine learning to find the most accurate algorithms that predicted future admissions trends.

The result is reduced waiting times for patients and better quality of care.

Product Development

With big data techniques, R&D teams are able to find useful data much faster and more efficiently, therefore reducing the time needed to develop the product and get it to market.

Healthcare Fraud Detection

Healthcare Fraud Detection

The common cases of fraud in healthcare include the following:

- Illegal medical billing practices in which claims are falsified.
- Multiple claims are filed by different providers for the same patient.
- Patient identities are stolen and used to gain reimbursement for medical services never provided.

Predictive Analytics for Fraud Detection

- Predictive analytics identifies patterns that are potentially fraudulent and then develops sets of “rules” to “flag” certain claims.
- These process is repeated in order to identify more and more emerging fraudulent patterns and create new “rules” for those as well
- Finally the best models not only flag the potentials but provide the reasons for that flagging, so that investigations and assessments by management can be completed efficiently.

Link Analysis for Fraud Detection

This method is used to detect the relationships among people, providers, and claims.

The auto insurance industry uses it to uncover fraud rings involved in “crash-for-cash” schemes.

For fraud protection in healthcare claims, algorithms can evaluate connections between individuals, providers, employees, and even suppliers of equipment, supplies, and pharmaceuticals.

Risk Scoring for Fraud Detection

This is a predictive model that can evaluate healthcare providers based on analyzing their claims, billing, and other pertinent data.

There can be any number of variables built in, and scores can be attributed based on those variables. Usually, the higher the score the greater the risk, and beyond a certain score, a provider can be flagged for an analyst to dig deeper and look at scores of individual variables. Think of it as similar to a credit score, based on several criteria.

Cluster Analysis for Fraud Detection

This type of predictive modeling puts providers, for example, into clusters, based on predetermined criteria, or parameters. They can then be compared to one another to see patterns that are normal or those that veer from that normality. Analysts can determine custom and unique clusters dependent upon payers' needs, and use those needs as a guide to create the right algorithms.

THANK YOU