

Data Analytics and Optimization



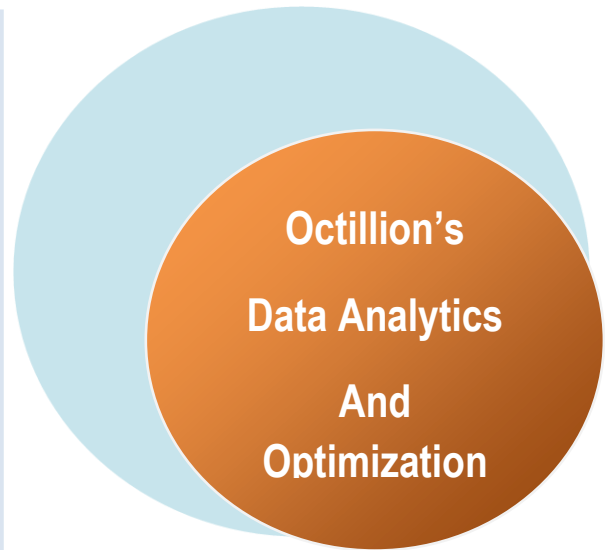
Contents:

1. Recommendation Engine
 - a. 360 degrees view of customer data
 - b. SmartMart Tools
2. Proactive Infrastructure Management
 - a. Infrastructure Management Framework
3. Knowledge Discovery from Unstructured Text Data
4. Executing Business rules in real-time

Data Analytics and Optimization

WHAT WE DO...A quick glance

- **Analyze:** Transactional and secondary data to revolutionize the usability, process and decision making
- **Design:** Enterprise 2.0 components to generate the required secondary data
- **Build:** Decision support systems that help all the users (not just decision makers) with real time predictive support
- **Create:** Expert systems that constantly optimize the software and help it perform better with the knowledge gained from each new transaction



"Fostering Smart Solutions"

PEOPLE

Octillion's data analytics team has a right mix of highly qualified team (Ph.Ds in engineering, computer science, Artificial Intelligence, Mathematics and Statistics) and works closely with the Artificial Intelligence teams from Carnegie Mellon University, IIT Hyderabad and Montclair State University.

WORK

Octillion's data analytics and optimization solutions help you identify and predict probable risks at various stages of your business processes and guide you to take proactive decisions.

So far, Octillion team has filed 5 patents and 3 more patent applications are pending in the area of Telecom network optimization, Customer churn and Shopping cart design. Our engineers and scientists regularly publish in international journals.

In addition to the cutting edge R&D, our teams are involved in solving real world problems in various verticals including retail, telecom, financial and healthcare domains.

BELIEF

Analytics is not just about using tools and techniques but a host of proactive approaches. The research and analysis to accomplish fact-based insights provides the intelligence required for today's smart decision making.

Irrespective of any business sector, the entire process of reaching to an optimal decision is a journey where facts, past experience, constraints and available resources are analyzed.

Data Analytics and Optimization Recommendation Engine

Customer analytics is a vital part of retail applications that provides the foundation for insights and recommendations on consumer purchase patterns. To cope with the changing landscape of customers' increased expectations and their buying behaviors, there is a need for most retailers to INTEGRATE data from multiple sources (POS, CRM, merchandizing, etc.), ANALYSE and present the analytical insights in REAL TIME.

Octillion's data analytics will help clients in seamless integration of data from disparate sources, perform analysis and produce real time decision support. With an objective to provide a holistic view of consumer behavioral data to the clients, our team has designed and patented the retail framework with the components mentioned below.

- 1) Transaction Scoring
- 2) Product Scoring
- 3) Customer Scoring
- 4) Manager Scoring (Competitor Score)

The system supports the following business decisions to answer some of the questions mentioned below:

- a) What product can be given as discount?
- b) Whom to give discounts?
- c) Which customer to be given what % of discount?

The above mentioned questions are examples and real life usages of these scores can vary from client to client. Each of these patent components can also be used as standalone depending on client requirement.

A 360 degrees view of customer specific data

Transaction Scoring:

The system uses transactional data of customers and runs market basket analysis to figure out which product goes well with what other product.

Based on the affinity rules generated, the system will also suggest the most common URL's from which the product was selected.

What can be achieved?

- Transaction based score rather than individual purchases
- Both application navigation and transactions are analyzed

Product Scoring:

Product scoring is computed based on the frequency of purchase, average shelf life, expiry date of the product, margin and supply-chain from vendors. Also, consumer rating of the product based on quality, price and brand will be taken into account in the computation of product scoring. This will help to decide the discounts that can be given on products.

What can be achieved?

- Product parameters
- Textual mining of user reviews

Customer Scoring

The system will score customer based on four components namely, Statistics, Segmentation, Network score and Churn score. The statistics based on demographic details and Market segment like the RFM score and monetary score. Network score based on how many times the customer has referred and finally Churn score.

What can be achieved?

- Churn probability
- Segmentation
- Networking & non-transactional activities

Manager Scoring

One of the powerful ways of understanding retail business is to understand competitor strategy. This module works like a dashboard to help managers understand competitor pricing. This main objective of this module is to recommend price based on competitor's price for top performing products.

What can be achieved?

- Competition scoring
- Seasonality scoring
- Business rules

Data Analytics and Optimization SmartMart Tools

Three other powerful components of recommendation framework are the Real-time Discounts, Gift Baskets and Campaign guide.

Real-time Discounts:

The primary purpose of this module is to decide how much discount can be given to a particular customer based on three parameters namely RFM segment, Network segment and Churn segment. All three are described above.

Retailer already knows how much the discount percentage for a product is. What is more important from a retailers' perspective is to know how much is the discount percentage for a particular product for a particular customer. This module facilitates the discount procedure in a more optimal way.



Gift Baskets:



This module innovatively applies the affinity rules discovered from Market Basket Analysis to design a customized gift basket. It can also be helpful in finalizing customer specific deals in optimal way.

This module takes input from RFM segment, churn segment and network segment. The objective of this module is to display a list of customer who satisfies certain criteria pertaining to each of the segment.

Campaign Guide



Data Analytics and Optimization

Proactive Infrastructure Management

Octillion pioneers Proactive Infrastructure management which is different from the conventional Reactive Infrastructure management approach. In the later, the emphasis is on providing in-depth, user friendly reports to the IT teams to enable them to manage effectively. Also, the only goal is to check whether a particular application or a component is working well or not and if something is not working well, it is brought to the notice of concerned person.

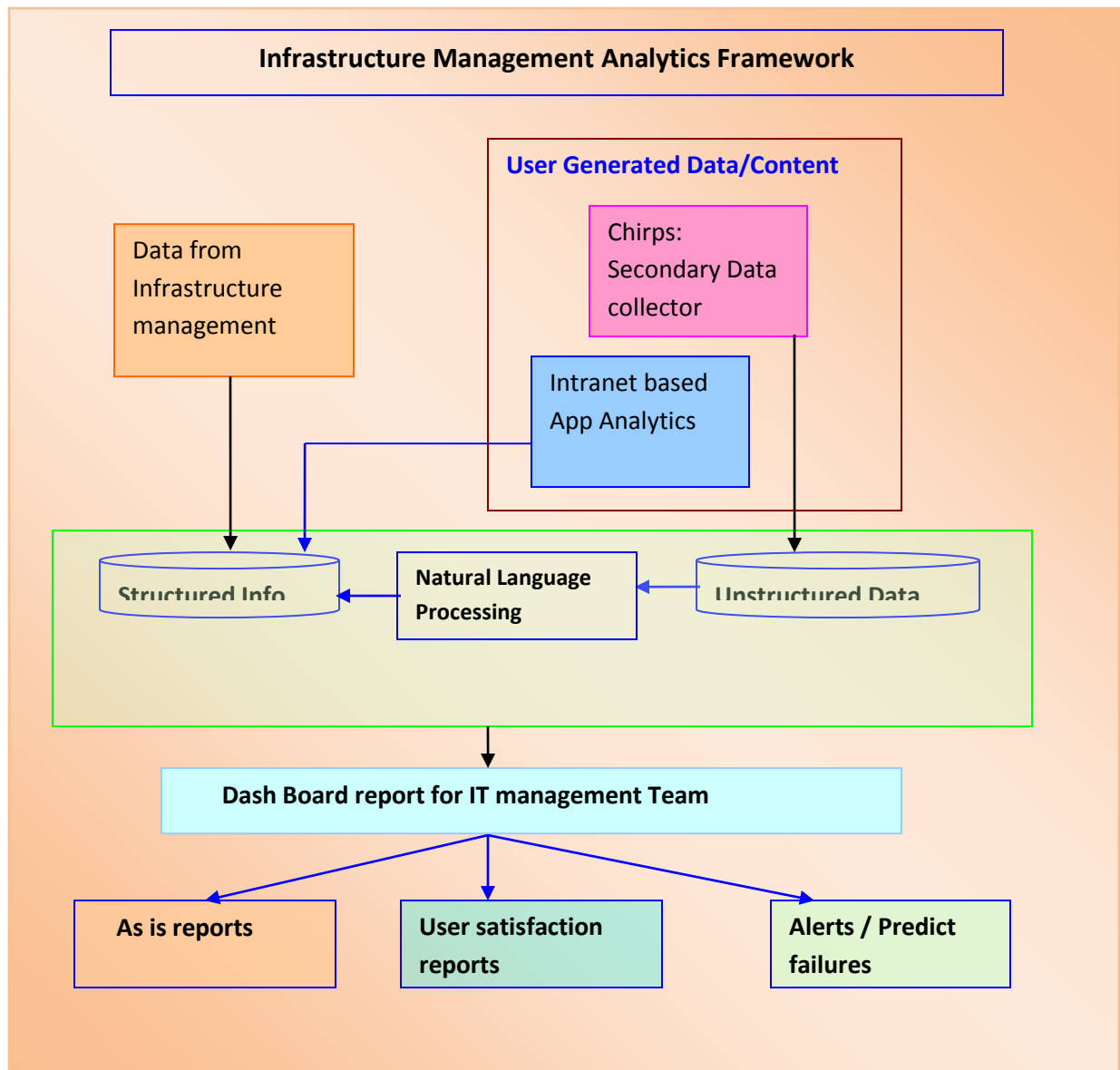
In the former, there are two additional goals that bring a sea of change to the infrastructure management

1. Applications/equipment don't fail all of a sudden. So, by monitoring the metrics pertinent to an application and historically correlating them with failures, ***we can define a set of statistical rules that can predict a failure ahead of its occurrence with a certain probability.*** This helps the IT team provide holistic cure of the problem rather than fire fighting
2. A well-functioning equipment/application does not mean users are happy using it. It just means it is available for people to use it. By combining the information obtained from application analytics (usage patterns), people's feedback and the uptime, ***we can measure how effectively people are using each page of the application.*** This helps the management constantly update their applications so they become user friendly and the users in turn become more productive

Proactive infrastructure management, hence, is about *attention and prediction*. Using the secondary data, Octillion answers the following important questions that are not typically addressed by any existing software:

Feature	Existing services	Octillion value
A consolidated view of the infrastructure	Yes	Yes
Real time alerts	Yes	Yes
Prediction of failures	No	Yes
Usage patterns and analytics of the apps	No	Yes
Satisfaction levels of the users for each application	No	Yes

Data Analytics and Optimization



Data Analytics and Optimization

Knowledge Discovery from Unstructured Text Data

THE SCIENCE BEHIND OUR TECHNOLOGY

Octillion uses a range of techniques, such as linear regression, neural networks, support vector machines, time series techniques etc for solving these problems. There is no single technique that is good for all problems, as each of the methods have their unique advantages in comparison with others. For example, if it is necessary to build relationship models between the input features and outcomes, regression model is a better choice than neural network. However, if accuracy were more important, then neural network would be the preferred choice. Depending on the type of the data available and the problem definition the best suited technique is used.

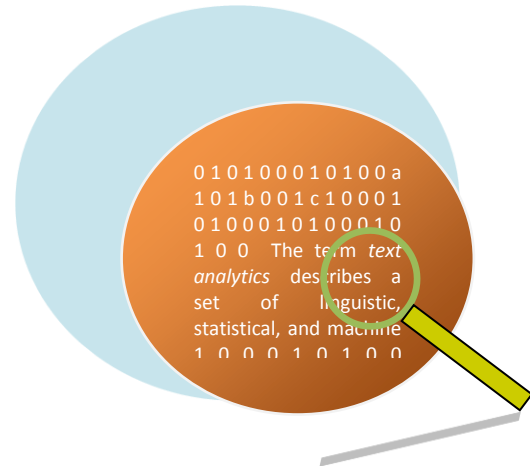
A brief description of some of the techniques used in Octillion

Linear Regression: This is a statistical approach to model the dependency of multiple variables to predict a single output variable. The underlying model is defined as a linear relationship between the inputs and output that fits a straight line to a cluster of points. The model is fairly simple, hence cannot fit complex relations with high accuracy. But it is the basis for other advanced techniques and can perform really well in sparse data sets.

Moving average: In this approach a subset window of fixed size is moved on the complete dataset. The average evaluated in the each of the window bin is used in building the models. The objective of using a moving window is to smooth the data and remove noise. When the moving average is applied to time series data it is referred to as autoregressive moving average (ARMA). There are different variations of ARMA models based on the lag definitions and input variables used.

Support Vector Machine: SVM is a combination of multiple linear regressions, where the space is divided into hyper planes using multiple linear cuts. This can also be further extended to use kernel functions and provide nonlinear fit capabilities.

Artificial Neural Network: ANN is based on the mathematical models of biological neural networks and are capable of mapping complex nonlinear functions. The mean squared error is used as the learning function to adapt the weights of the neurons and provides a high accuracy in prediction on training data.



TEXT TO KNOWLEDGE CONVERSION

Automated Text Classifiers: Classification of texts into multiple category spaces is a major application if we need to categorize the uncategorized content. Based on machine learning methods (such as Support Vector Machines, Logistic Regression, Random Forests, Exponential Models, etc.), text may be automatically classified by topic (e.g. pertaining to cell phones, or to movies, or to cars, or to electronics), by genre (e.g. product reviews, vs. advertisements, vs. news items, vs. FAQs), and subcategorized by more specific criteria (e.g. by manufacturer).

Sentiment Analysis: Combining machine learning and rule-based methods, the writer's sentiment can be automatically determined, e.g. highly favorable, ambivalent, mildly positive, outraged, etc. If multiple products or services are mentioned in the same text, performing the attribution of sentiment to the correct products requires some further development, but is within scope. Our methods are robust to linguistic structure, classifying "I would rather use the great product from the competition instead of this poor excuse for a ..." on a product review as *strongly-negative*, rather than be tricked by the presence of "great" or "great product" and classifying it as *positive* or *strongly positive* by all commercial systems we tested.

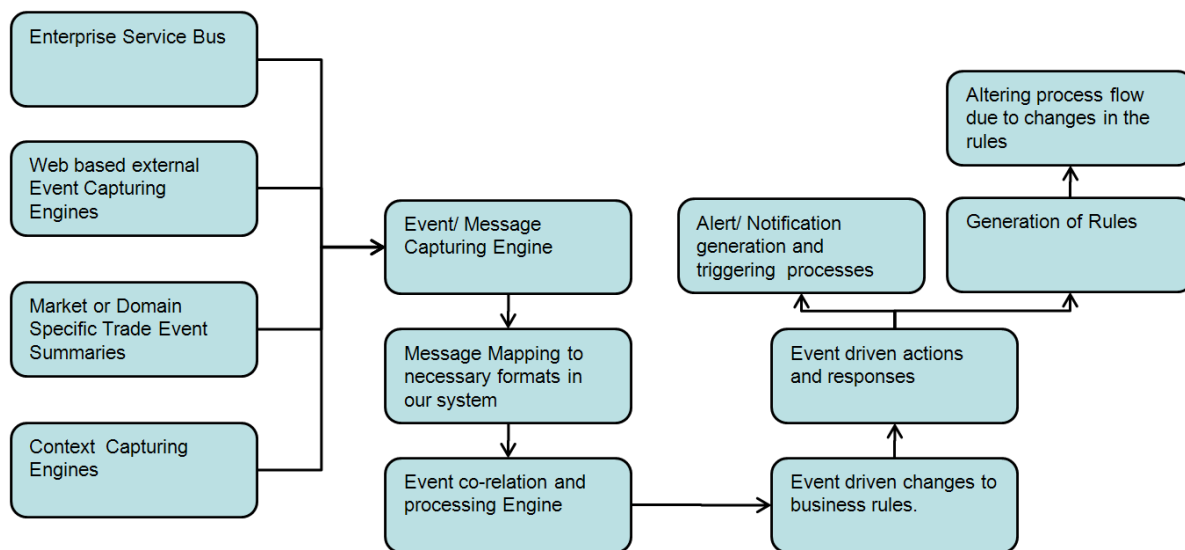
Search Engines: Our partner university CMU developed "Lemur and Indri", two advanced search engines. Both apply to virtually any type text documents and web pages, and Indri also combines structured search on metadata (e.g. author, date, subject codes, access/security codes, etc.). These are open-source and we have well trained resources to integrate these

Data Analytics and Optimization

Real-Time Decision Making with Auto Generated Business Rules

Octillion’s solution for real time decision making analyzes the isolated and seemingly unrelated business and system events, this analysis is translated into business rules that can be embedded into operational BPM, workflow, alerting and responsive actions.

The overview of the system is depicted below.



Going one step below, seemingly disparate processes can affect each other significantly. Consider this scenario: In the airline industry, any time a disruption happens, immediate adjustments must be made to ensure that equipment, crews, passengers, and baggage all get where they need to be. This process requires the interaction of many systems and in some cases, human intervention, leaving room for error and creating inefficiencies. What if the airline operations team can Gain real-time visibility into flight operations, anticipate the downstream impact of any disruption and Continuously improve recovery processes based on the new rules being generated by our solution. The rules could typically change the ability for users to book new tickets on the airline or help reduce traffic from other connecting flights to accommodate this sudden alteration of traffic.

Octillion’s frameworks collect data from multiple sources, (accepts both structured and unstructured data) and continuously churns out business rules/ alters existing rules. We not only provide preliminary predictive analysis, but also provide advanced statistical

Data Analytics and Optimization

analysis using Artificial Intelligence techniques. For instance, for a large trucking company, we could help the users with:

- Directing and predicting more effectively material, trucks, people etc. to places they are required.
- Monitor accurate estimated time of arrivals (ETA) to docks and ports and automatically direct resources based on the rules.
- Provide alerts when rules determine something is not moving as expected
- Detect, in real-time, and automatically respond to supply chain ensuring optimal capacity usage

Our framework allows detection of data anomalies as the data flows in from multiple seemingly unrelated sources and churns out rules that can be embedded in the existing operational systems. This would be based on our systems ability to constantly learn from the previous anomalous patterns, its impacts and enhance real-time predictions.

Below is the process that depicts the traditional versus Octillion’s implementation of real-time analysis of anomalies in data.

