



# Octillion's Data Analytics and Optimization

*"Fostering Smart Solutions"*

Octillion Technology provides custom IT solutions to solve complex business problems. Our expertise is in prediction, risk analysis and optimization. Our Carnegie Mellon University (CMU) trained analysts deliver best value using state of the art technology.

Octillion's Data Analytics and optimization capabilities will empower you to take intelligent decisions and achieve superior results.

Octillion offers a wide range of *"Smart"* solutions in forecasting, prediction, bench marking and unstructured data to structured knowledge conversion.

## PEOPLE

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

Our technology team has filed 5 patents. Additionally, 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.

## 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.

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

## BELIEF

Analytics is not just about using tools and techniques. It involves 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 business vertical, the entire process of reaching to an optimal decision is a journey where facts, past experience, constraints and available resources are analyzed.

## 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 has 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.

## KNOWLEDGE DISCOVERY FROM UNSTRUCTURED DATA

---

0 1 0 1 0 0 0 1 0 1 0 0 a 1 0 1 b 0 0 1 c  
1 0 0 0 1 0 1 0 0 0 1 0 1 0 0 0 1 0 1 0 0  
The term *text analytics* describes a set  
of linguistic, statistical, and machine 1  
0 0 0 1 0 1 0 0 learning techniques that

---

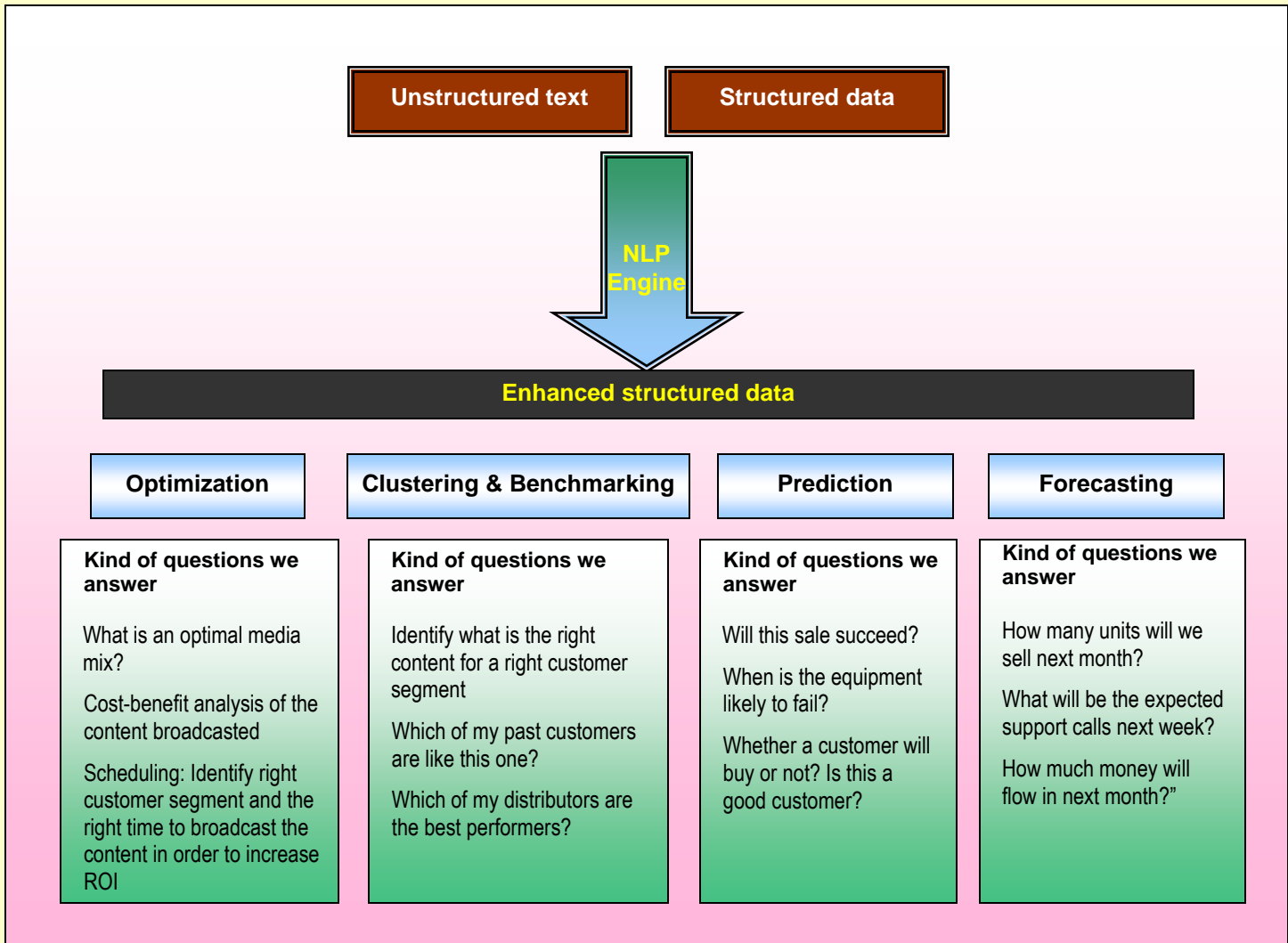
## TEXT TO KNOWLEDGE CONVERSION

**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.

Some of the questions that Octillion’s Data Analytics and Optimization solutions answer are listed (but not limited to) below. Based on the client specific need, we either provide a complete solution where we collect the unstructured information along with the structured data or a structure data only to formulate and develop models.



- QUICK GLANCE....**
- Wide range of smart solutions in forecasting, prediction and bench-marking
  - Strong team of highly qualified and experienced engineers and scientists
  - Advanced technology solutions such as Natural Language Processing to convert text data into usable structured information
  - On-site and off-shore availability of technology implementation and support services

**Corporate HQ**  
**Octillion Technology Inc.**  
 (www.OctillionTechnology.com) is a C corporation.  
 Our offices are located at 82 Pioneer Way, #105,  
 Mountain View, CA 94041  
 USA  
 Please call us at (866) 429 9532 or write to us:  
 info@octilliontechnology.com

**India Operations**  
 Octillion Technology Pvt Ltd, is a private limited company. Our office is located at:  
 # 2-56/2/19, Khanamet Madhapur,  
 Hyderabad - 500081 India  
 Please call us at (866) 429 9532 or write to us:  
 info@octilliontechnology.com