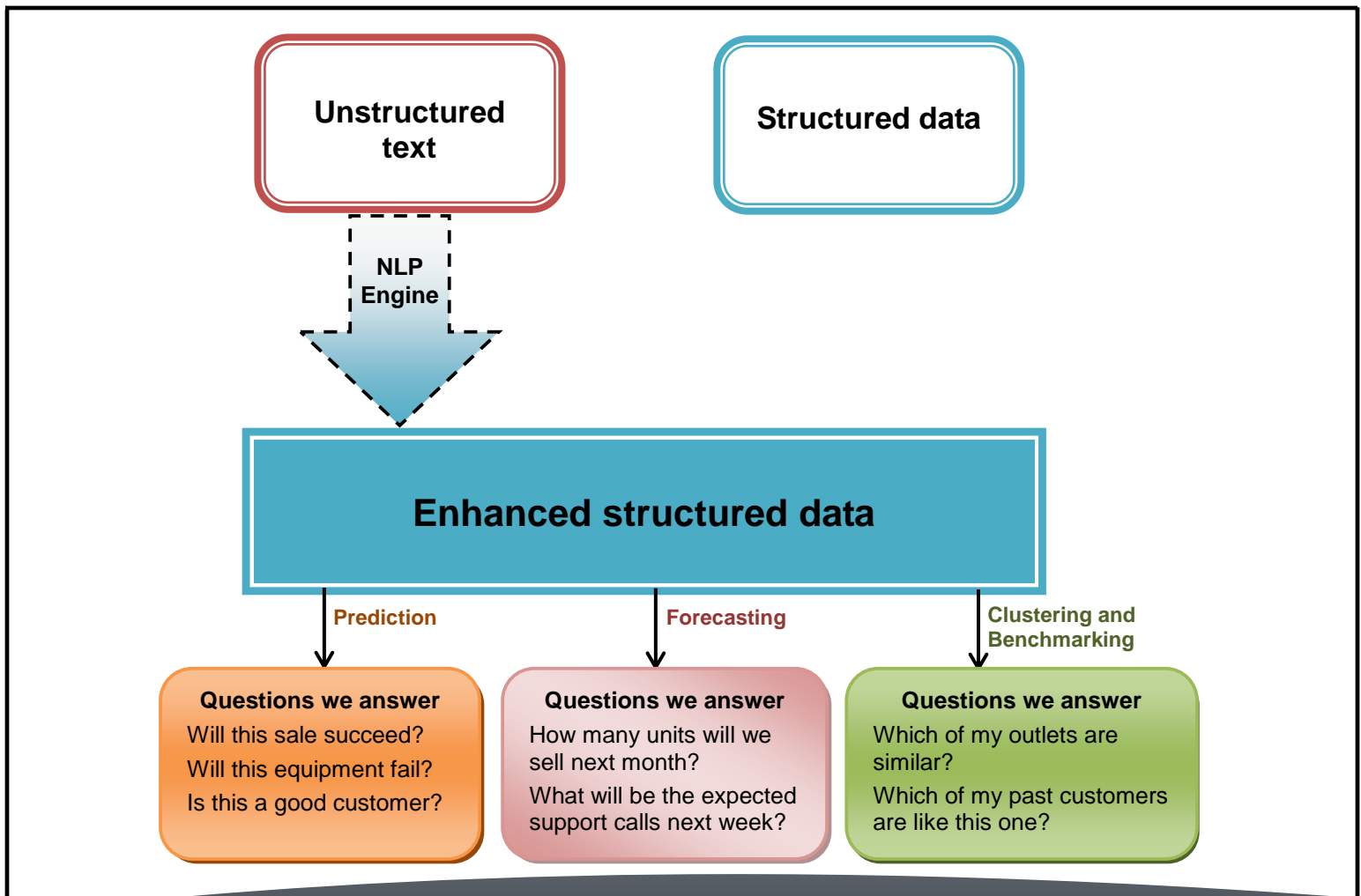


Introduction

Octillion Technology offers a wide range of “Smart” solutions in forecasting, prediction, bench marking and knowledge conversion of unstructured data to structured knowledge. Some of the questions we help our clients answer accurately are

- (a) **Prediction:** “whether a customer will buy or not?”, “Whether an equipment will fail or not” or “whether a sale will be successful or not”
- (b) **Forecasting:** “How many items will be sold” or “How much money will flow in next month?”
- (c) **Benchmarking:** “which of my sales people are the best performers” and “which of my outlets are similar”
- (d) **Data2Knowledge:** In addition, we specialize in converting textual information in public domains or in any legacy application to tagged structured data. Once the conversion is done, the sentiment detection and most other BI queries can be easily executed on this structured data

UNSTRUCTURED DATA TO STRUCTURED KNOWLEDGE CONVERSION



Our solutions find unique applications in knowledge management, resource estimation, risk management and future planning.

In general, the techniques utilized for building forecasting models require large amounts of data and are computationally intensive. With the recent advancements in high performance computing and large data visibility, the adoption of forecasting techniques to everyday business problems has become an industry best practice.

Octillion Technology solutions use 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 its 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.***

Enclosed is a brief description of some of the techniques used in our solutions:

1. 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.



- 2. 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.
- 3. 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.
- 4. Artificial Neural Network:** ANN is based on the mathematical models of biological neural networks and is 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.

In **text to knowledge** conversion, we specialize in the following

- **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



About the Team: Octillion Technology analytics team has a strong internal team (PhDs in engineering, computer science, AI, Mathematics and Statistics) and works closely with the AI teams from Carnegie Mellon University. The team has filed 5 patents last year and 3 more are pending. Our team has published in several international journals. In addition to the cutting edge R&D, it's involved in solving real world problems in various verticals including retail, engineering, telecom, financial and healthcare domains.