Computaris

Caring - Innovating - Transforming

Deciphering Voice of Customer through Speech Analytics

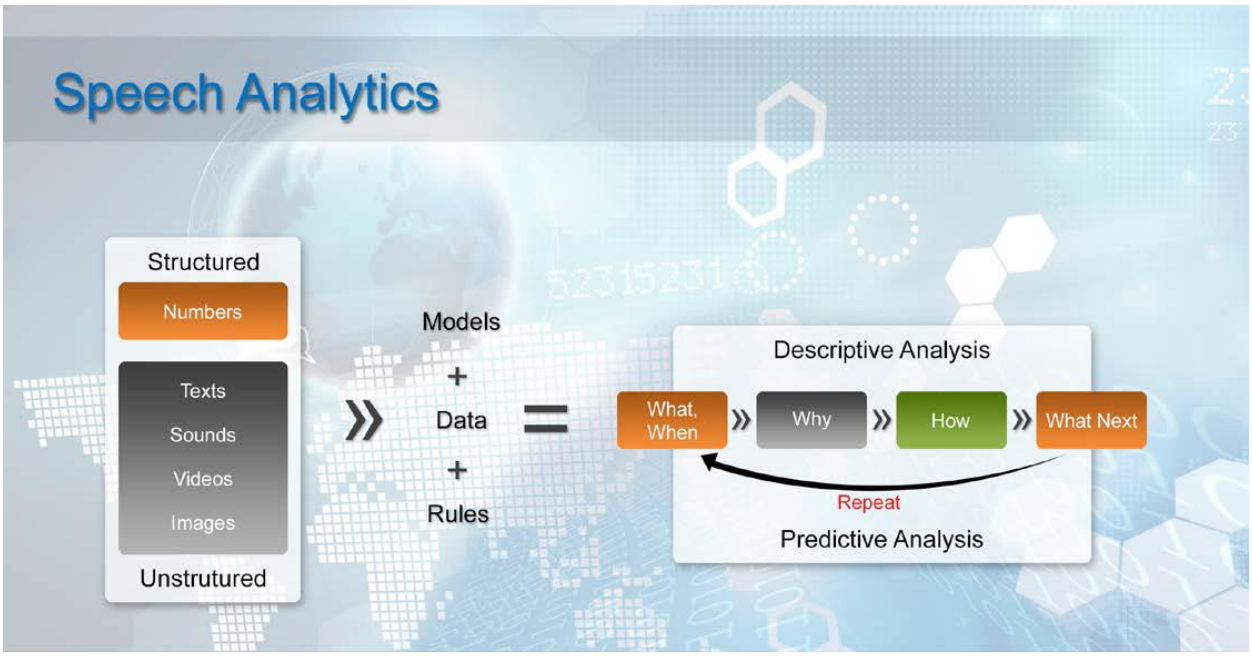
Whitepaper

Computaris
Caring - Innovating - Transforming

Contents

1 INTRODUCTION

- For most retail businesses, customer interaction via call centers is a very significant communication channel. Organizations typically receive thousands of customer calls every day. According to an industry report, over 56 million hours of conversations (nearly 420 billion words) are spoken a day in call centers worldwide. If the audio data thus collected can be aggregated and analyzed, it can yield quality insights into customer expectations, preferences, service issues & product usage. While speech analytics is not a new technology to the market, most of the business executives are still skeptical about the value it can add.
- This whitepaper aims to illustrate basic technologies used in speech analytics, their use cases and how RoI from speech analytics software can be maximized.



2 WHAT IS SPEECH ANALYTICS?

Speech analytics is a powerful tool for analyzing recorded calls, structuring customer interactions and gaining insight in the hidden information. It can be used for audio mining, speech categorization, intelligence extraction, decision making, monitoring agent performance.

If applied correctly and used effectively speech analytics can help improve service quality, reduce operating expenses, boost revenue, and reduce customer attrition. If integrated well with overall strategy it can help businesses drive product & process innovation leading to significant market differentiation. However there are significant challenges in transforming speech data to a structured form which can be subjected to further analysis.

2.1 Challenges analyzing speech data

- The speakers differ in speaking style and speed, gender, age, dialects, physical attributes (such as vocal tract). Any speech recognition system has to take all these features into consideration. For example “service provider” may be recognized as “serve his provide her”.
- Humans in addition to speech, also communicate via facial expressions, emotions, postures and eye movements, and these are missed by an automatic recognition system (ASR).
- While interacting in the real time environment humans encounter lot of unwanted sounds called noise and these need to be filtered from the speech signals.
- Homophones (i.e. words that are pronounced the same but differ in meaning. For e.g. two & too) and word boundary ambiguities pose a major problem to speech recognition systems.
- The acoustic waves change with the properties of the channels used.
- The continuity in speech leads to problems related to identification of word boundaries.
- Grammatically spoken language is very different from written languages.

2.2 Algorithms for Transforming Speech to Structured Data

To fully leverage the information encapsulated in customer calls, you need to transform the interaction data in the audio files to a more structured data which can be queried by analysts and can be consumed by sophisticated machine learning algorithms.

Figure 1 gives the complete high-level flow of the process for conversion of unstructured voice data to a structured and more useful information.

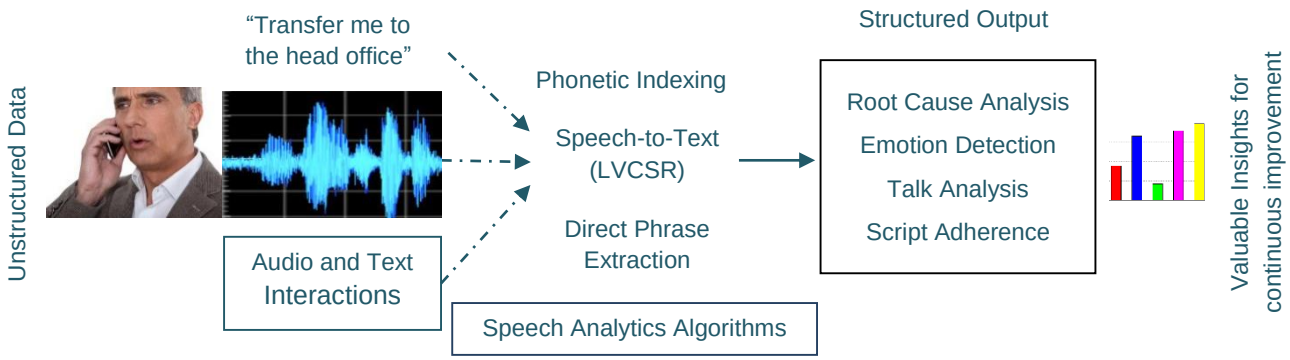


Figure 1: Flow for transformation of unstructured to structured data in speech analytics

The different speech analytics approaches such as phonetic indexing, speech-to-Text (LVCSR) and direct phrase extraction achieve this by deploying a speech recognition system which comprises of some or all of the following components: an acoustic model, grammars, a language model and recognition algorithms. These components are standard in any speech analytics software today. But understanding them is important as the quality of speech analytics is dependent on how these components are configured and what algorithms they use.

2.3 Key Components of a Speech Recognition system for Analytics

Acoustic model: an acoustic model represents the relationship between an audio signal and the phonemes or other linguistic units that make up speech. An acoustic model contains statistical representations of each of the distinct sounds that makes up a word. An acoustic model is created using a speech corpus and training algorithms that create the statistical representations called Hidden Markov Models (HMMs), for each phoneme in a language. Each phoneme has its own HMM, and HMM is one most common type of acoustic models. The different components of a speech recognition system are shown in Figure 2.

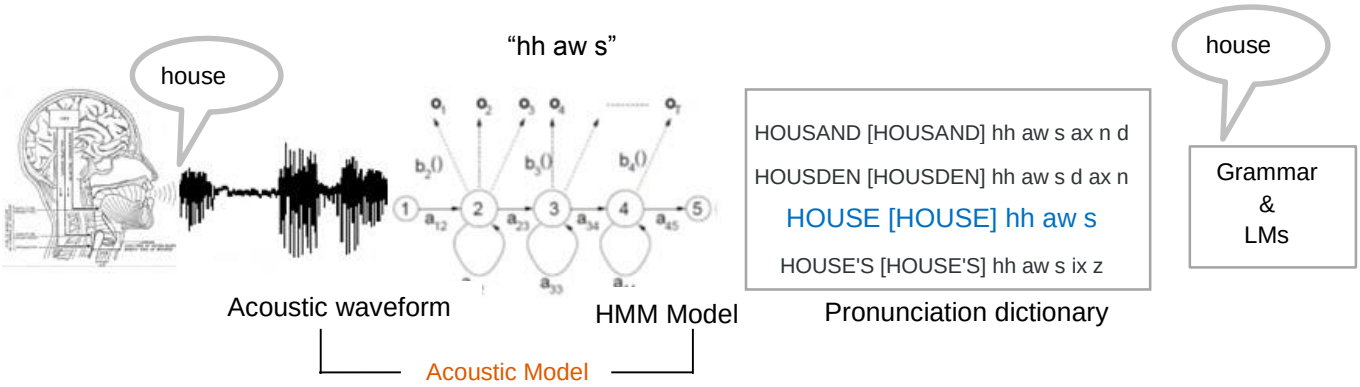


Figure 2: Different components of a Speech Recognition System

Language-specific *acoustic models* are used directly in the **Phonetics indexing approach** being used for speech analytics. The basic recognition unit for this approach is a phoneme and it is a dictionary

independent approach. It allows users to query phonetic strings and perform orthographic search using a pronunciation model.

Language model: a statistical language model (SLM) consists of a list of words with their probability of occurrence. It is used to restrict search by limiting the number of possible words that need to be considered at any one point in the search. This results in faster execution and higher accuracy. Tri-grams are the most commonly used LMs in ASR. The probabilities of n-grams help in determining which n-gram is more probable in comparison to the other similar n-grams. For example the P (I saw a van) >> P (eyes awe of an) for correct recognition of the phrase. This is calculated using a language model similar to what is shown below.

P(I|<s>) = 0.67	P(eyes|<s>)=0.25	P(saw|I) = 0.63	P(awe|eyes)=0.33
P(a|saw)=0.5	P(of|awe)=0.45	P(</s>|van)=0.6	P(</s>|an)=0.15

The large-vocabulary continuous speech recognition (**LVCSR**, also known as speech-to-text (STT) or word spotting) speech analytics approach uses both language-specific *acoustic models* and *language models*. The basic unit for this approach is a set of words, and these are generally bi-grams or tri-grams. In order to map words to phonetic forms it also uses a pronunciation model or a dictionary. This is a dictionary dependent approach.

Recognition algorithms: perform speech recognition based on written grammar. This grammar describes the possible patterns of words. The recognition grammar is generally given using two files the 'grammar' and the 'vocabulary' file. The 'grammar' file defines category-level syntax, whereas, the 'vocabulary' file defines word candidates in each category, with its pronunciation information. For illustration, consider the sentence "I'll take one house please". The 'house.grammar' in the BNF would be as shown below. 'S' indicates the sentence start symbol. The rewrite rules are defined using ':' symbol.

```
S : NS_B HMM SENT NS_E
S : NS_B SENT NS_E
SENT: TAKE_V HOUSE PLEASE
SENT: TAKE_V HOUSE
SENT: HOUSE PLEASE
SENT: HOUSE
FRUIT: NUM HOUSE_N
FRUIT: HOUSE_N_1
```

The vocabulary file contains definition of each word defined in the 'grammar' file. The partial 'house.voca' file for the 'house.grammar' file is given below.

```
% NS_B                                   % NS_E                                   % HMM
<s>             sil                     </s>            sil     FILLER          f m
                                                                FILLER          w eh l
% TAKE_V                                 % PLEASE
I'll take       ay l t ey k             please          p l iy z

% HOUSE_N                                % NUM
house           hh aw s                 one             w ah n
                                        two             t uw
%HOUSE_N_1
Houses          hh aw si z
```

2.4 The Technology Comparison

There has been a long-standing debate on the merits of these approaches. Several factors are considered while choosing the technology. There is no one best approach that may be pointed out, however, the approach may be chosen based on the requirements as discussed below.

- *Targeted listening or calls of interest*: If you aim to listen to calls containing specific keywords, then a system with dictionary dependent (LVCSR) may be preferable. The dictionary dependent approaches can recognize words that are already in the lexicon.
- *Out of Vocabulary (OOV) words*: If it is very likely that you encounter new words in your search domain, then dictionary independent approach should be preferred. OOV handling is a major issue in the dictionary dependent approaches.
- *Audio Processing*: The audio needs to be reprocessed for new words added to the dictionary. This proves to be a time consuming and a costly process, as experts are also needed for entering words in the dictionary. The audio is processed just once in the dictionary independent approaches.
- *Speech transcripts*: The dictionary independent approaches (phonetic), due to absence of a language model cannot be used to generate a meaningful orthographic transcript of speech.
- *Precision vs. Recall*: A higher precision rate may be obtained with a dictionary dependent system on words that are already in the dictionary, but the recall rate suffers as there are always some missing words. This also results in higher error rates. Phonetic approaches have low precision but higher recall.

Features→ Approach↓	Speed	Recall	Precision	Error Rate	Reliability
Phonetics	High	High	Low	Low	High
LVCSR	Low	Low	High	High	Low

Table 1: Comparison of the approaches

3 ORGANIZATIONAL USE CASES FOR SPEECH ANALYTICS

Some of the use cases of speech analytics in a call center could be:

- **Tone and Sentiment Analysis**: Speech analytics systems can analyze the tone and detect the sentiment of voice. Tone can also signify age, this can be used to determine the effectiveness of a marketing campaign on a specific age segment.
- **Talk and Silence Pattern Analysis**: The talk and the silence patterns can be analyzed to measure emotions and levels of satisfaction. A set of user-defined phrases can be used for detecting agent actions. Thus, it helps in identifying and prioritizing what needs immediate attention.
- **Agents' Performance Monitoring**: Monitoring agents' interactions with customers can easily detect proactive agents and also agents not successful in satisfying customers. This analysis may be used for training the agents for improving their performance.
- **Call Segmentation**: Certain calls may be difficult for agents to handle. Such calls may be identified, segmented, and implemented using specific business processes.
- **Decision making**: The insights gained from the information may help in making decisions and implementing new policies for product & service improvement.

4 SOLUTION LANDSCAPE: Vendors, Products and their Market Share

According to DMG consulting and ContactBabel[1], only 24% of the organizations are currently using a speech analytics solution. However, interest in speech analytics is growing and the market for this will continue to expand over the next several years. DMG consulting estimates this growth to be 18% in 2015, and 16% in 2016 and 2017. The market shares of some of the leaders of speech analytics organizations is summarized in Table 2.

Vendor	Product	Approach used	Market share1(%)
NICE	NICE Interaction Analytics	Phonetic indexing and Transcription	30
Verint	Impact 360 Speech Analytics	Audio indexing	24.5
HP Autonomy	HP IDOL (Intelligent Data Operating Layer	Meaning based computing (MBC). MBC stresses relevance along with accuracy	13.1
Nexidia	Neural Phonetic Speech Analytics	Automatic speech recognition (ASR), phonetic indexing and word-level transcription	8.7
Genesys (UTOPY)	Speech and Text Analytics (formerly, SpeechMiner)	Direct phrase detection and Transcription	4.2

Table 2: The product, approach & the market share of major vendors in speech analytics

Some other speech analytic vendors with their market share (%) are CallMiner (10.7%), Avaya (1.7%), Mattersight (1.7%), Calabrio (0.6%), and Interactive Intelligence (0.5%). An overall market share of top ten vendors is shown in Figure 3.

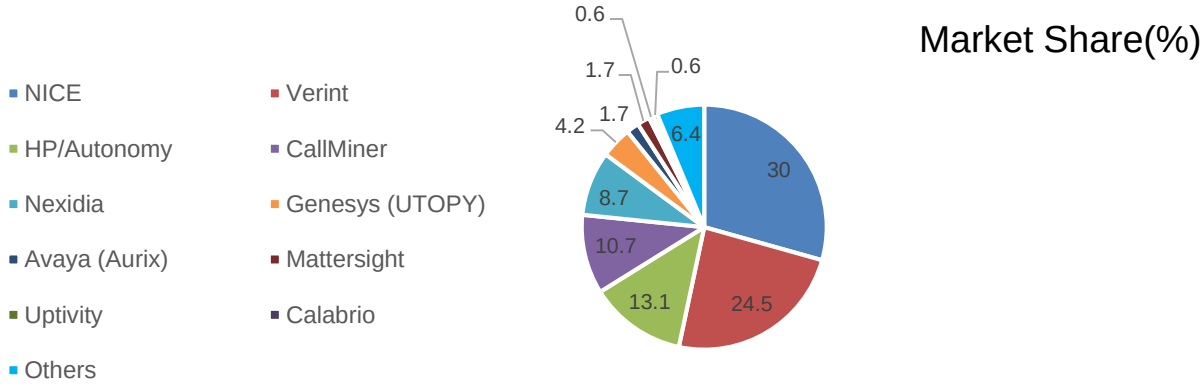


Figure 3: Top ten vendors' market share

[1] Speech Analytics Product and Market Report Reprint - DMG Consulting LLC (as of May, 2014)

5 RECOMMENDED ARCHITECTURE FOR SPEECH ANALYTICS

We propose a 2-phase architecture shown in Figure 4, which allows data discovery & predictive analytics on the voice call data. This architecture aims at classifying contexts and issues being constantly talked about on calls and predicting the customer behavior. Since the insights generated are contextual these can help product managers, service units and other units to derive strategic inputs. The R Analytics Engine (RAE) takes phonetics or text as input. The engine analyzes the input and models it based on the key factors, which proves helpful in deriving customer segment based strategic insights and predicting the KPI's and the behavior.

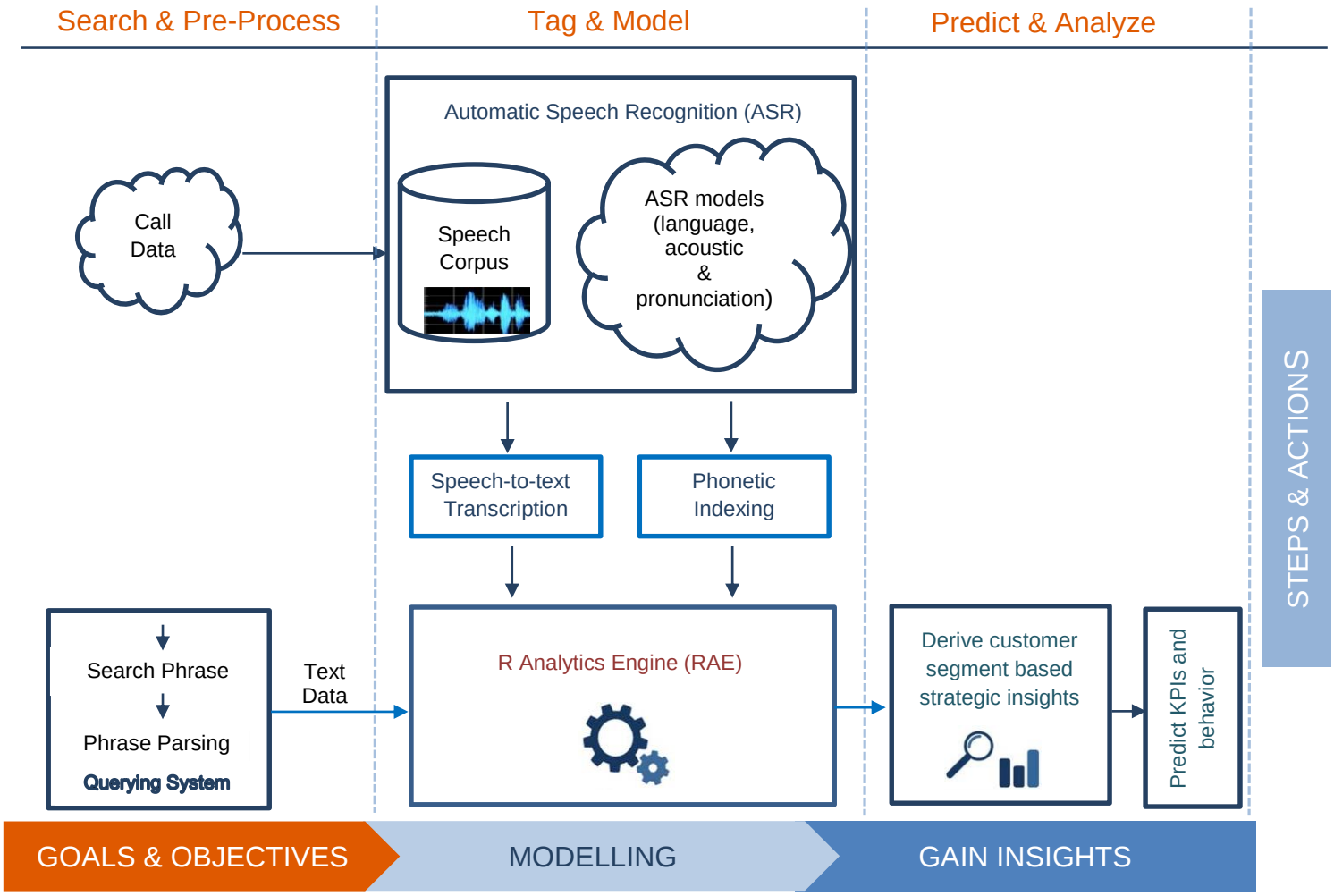


Figure 4: Recommended Architecture for Speech Analytics

6 CASE STUDY: Predicting NPS in Health Insurance

Net Promoter Score (NPS) is a metric frequently being used to measure an organizations' performance from customers' perspective, and to measure the loyalty existing between two entities. The entity could be a company-customer or an employer-employee. NPS is calculated based on the score provided by a customer on a scale of 0-10 for the question "How likely is it that you would recommend my company to someone?" This scale segregates customers into three categories based on the scores provided; the Promoters (9-10), the Detractors (0-6) and the Passives (7-8)[2].

NPS is computed as the difference between the % of Promoters and the Detractors on the basis of sample surveys which are conducted by organizations periodically. Predicting NPS on the basis of customer call & speech data is one of the most compelling use cases of speech analytics. Rather than waiting for next survey, speech analytics can be used to predict probable promoters and detractors which can help organizations estimate their overall NPS score in real time basis. In case of declining NPS score thus predicted, strategic changes can be done in business and service to maintain the desired results without waiting for next survey results.

Objective: Our objective of this case study was to predict promoters, detractors and passives from the call data available from a large health insurance provider in US. The data available included customer's behavioral data from CRM, the call characteristics (hold times, FCRs, recommended CSA etc.) and text transcripts generated using the speech analytics engine.

Challenges: The text transcripts as provided had lot of noise. The data had to be preprocessed and cleaned by our R Analytics engine to remove noise & convert it to a more structured form which was further amenable for analysis and modelling. We typically found 2-grams & 3-grams having more information gain ratios than uni-grams and hence better predictors of class categories.

Results: The impact of call category, call duration and other related factors was analyzed on the NPS score. It was found that the FCRs & the hold time had a significant impact on the NPS score. Addition of 2-gram & 3-grams extracted from text transcripts using speech analytics improved the NPS prediction significantly by 8.6%. We were also able to predict the NPS performance at agent, call center and campaign level. The model results were used to drive specific training agenda for call center agents.

[2] http://www.netpromoter.com/why-net-promoter/know

7 CONCLUSION

The information shared and exchanged through spoken interaction data largely remains untapped. Speech analytics can provide a solution by collecting this data and provide insight on these interactions. Speech analytics can prove to be a revolutionary approach in measuring the customers' emotions, context and the intent. Customers generally do not bother telling people about a great customer service experience but the same customer makes sure that almost everyone knows about a bad one. Therefore, business success depends heavily on the customer experience and so enhancing this experience is critical for success of any business. The R Analytics Engine (RAE) can help accelerate data discovery and mining of speech data.

About the authors

Renu Balyan is a Lead Analyst (Data Science) with R Systems International Ltd. Her areas of interest include data analytics, machine learning, machine translation and information extraction. She has published 14 papers in national and international conferences & journals. She is currently pursuing her PhD from IIT Delhi. She has worked as an intern with Dublin City University, Ireland. She has also worked as a research fellow and project engineer with Centre for Development of Advanced Computing, Noida & worked on various projects related to natural language processing (NLP) for nearly 6 years.

Praveen Pathak is the Analytics Practice Head for R Systems International Ltd. With over 14 years of consulting, in-house & offshore analytics delivery experience, Praveen focuses on providing R System's clients best in class analytics solutions and services. He has extensive hands-on as well as leadership experience in analytics, information management, predictive modelling, optimization and big data technologies to draw data driven insights and help address business challenges. His interests includes Artificial Intelligence, Neural Networks & High performance computing.

About Computaris

Computaris, an R Systems business, provides specialist BSS technical consultancy, software development and system integration services for the telecommunications industry in Europe, North America and South East Asia. Our company offers the highest level of expertise in the area of real-time rating and charging, messaging, provisioning, mediation, subscriber data management, mobile broadband data policy management, and loyalty and churn management. For more information, please visit www.computaris.com.

About R Systems

R Systems is a leading OPD and IT Services company, which caters to Fortune 1000, Government, and Mid-sized organizations, worldwide. The company is hailed as an industry leader with some of the world's highest quality standards, including SEI CMMI Level 5, PCMM Level 5, ISO 9001:2008, and ISO 27001:2005 certifications. With a rich legacy spread over two decades, the company generates value that helps organizations transcend to higher levels of efficiency and growth. Quite like the Oyster delivering the Pearl.

For more information, visit www.rsystems.com.

Got any questions?

Contact Computaris at:

Tel: +40 21 204 5242

Email: marketing@computaris.com