Cognizant 20-20 Insights

Digital Systems & Technology

Optimizing IT Operations with Natural Language Processing

As artificial intelligence becomes increasingly mainstream, natural language processing techniques are emerging to help IT teams gain enhanced understanding of their operations landscape and to further optimize the ticket management process.

Executive Summary

A substantial portion of any IT operation is taken up with maintenance and support of applications and infrastructure. As such, every problem or request is initiated as a ticket that is worked on manually by an operations team. In large operations, the volume of these tickets could run exceptionally high – thousands every month. Management of these tickets must be continuously optimized to keep operational costs under control.

Traditional approaches to such optimization require significant manual effort by subject

November 2019

matter experts (SMEs), which is a big inhibitor in the continuous improvement journey. SMEs usually categorize trouble tickets via a keyword-dictionary-based approach, but such dictionaries are difficult to maintain. Natural language processing (NLP) techniques help overcome this challenge, resulting in cost optimization as well as service level improvement.

This white paper lays out an NLP-based solution that assists SMEs in identifying patterns efficiently across thousands of tickets by assessing ticket descriptions automatically.

Challenges in identifying the right treatment strategies

Typically, end users describe their problem or request in free text as part of the ticket description. Although the range of issues is limited, as the underlying application portfolio has a defined functionality, users often describe the same or similar problems in many different ways.

To enhance efficiency, operations teams employ various strategies:

- Automating ticket resolution, which speeds and standardizes the whole process, from problem identification through closure.
- Documenting the steps required to resolve the tickets, known as standard operating procedures (SOPs).
- Assigning less experienced or right-skilled personnel to the tickets, a strategy known as "left shift."
- Eliminating the underlying problem and hence recurrence of such tickets, thus reducing ticket volumes.

To identify and deploy the right treatment strategy, SMEs must go through ticket descriptions and develop an understanding of the patterns of problems. Such an approach is effort-intensive and, when done manually, is often both inefficient and suboptimal.

Sometimes SMEs define a dictionary of keywords or phrases corresponding to different categories of tickets. To map tickets to categories, an automated program looks for the presence of these keywords in the ticket description. However, this approach has its own challenges:

- Such a keyword dictionary is never comprehensive and requires constant updates.
- The mere instance of a keyword is not sufficient for mapping tickets to categories.
- It allows mapping only to a set of predefined ticket categories.

To identify and deploy the right treatment strategy, SMEs must go through ticket descriptions and develop an understanding of the patterns of problems. Such an approach is effort-intensive and, when done manually, is often both inefficient and suboptimal.

NLP eliminates the need for SMEs to process individual ticket descriptions or define a comprehensive keyword dictionary. It also handles the diverse ways in which different users describe the same problem on their tickets.

Leveraging NLP to identify patterns in tickets

NLP can help address some of the challenges covered above by assessing the free text descriptions automatically across thousands of tickets. To accomplish this, a vector representation, or embedding,[1] of ticket descriptions is created using the word2vec[2] technique. The descriptions and other relevant ticket characteristics are used to identify patterns automatically and devise homogeneous groupings or clusters of tickets. Leveraging these insights, SMEs can not only identify the right treatment strategies but also develop an enhanced understanding of the ticket landscape.

NLP eliminates the need for SMEs to process individual ticket descriptions or define a comprehensive keyword dictionary. It also handles the diverse ways in which different users describe the same problem on their tickets.

Our NLP solution entails a three-step approach:

- Automated processing and clustering[3] of tickets.
- SMEs identify treatment opportunities and execute them – e.g., preparing an SOP.
- Assigning every new incoming ticket to the right cluster for appropriate processing.

Figure 1 illustrates how three similar tickets, articulated differently by the users, are grouped together under cluster number 2 without using any predefined keywords.

Automatic identification of semantically similar tickets

S. No.	Ticket Description	Assigned Cluster
1	SUP:SMP:Unable to launch ILT course — FAILED	2
...	...	...
...	...	...
102	SUP:SMP:user cannot finish a training	2
...	...	...
3013	SUP:SMP:Problem completing a web-based learning activity	2
...	...	...

Figure 1

Clusters identified in a given ticket repository

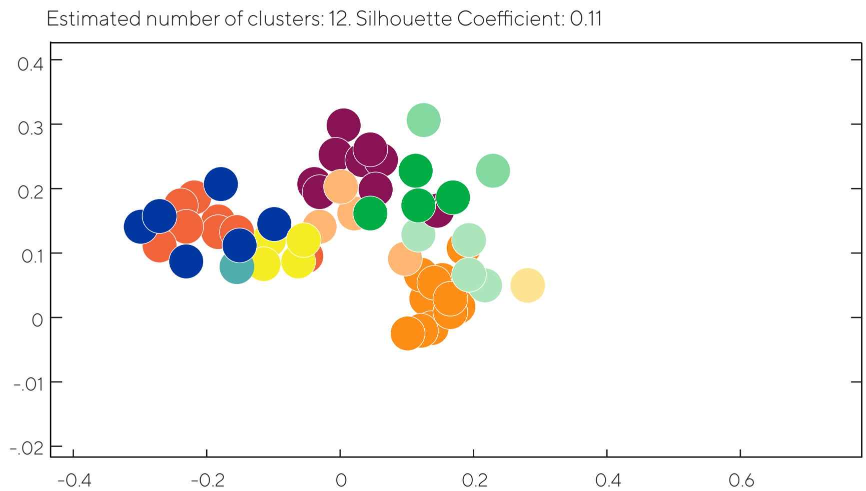


Figure 2

Figure 2 shows a scatter plot of tickets, color-coded by cluster, along with the optimal number of clusters suggested by the DBSCAN[4] algorithm and a measure of cluster quality called silhouette coefficient.[5] The plot axes are not directly meaningful and are used only to show cluster separation.

Figure 3 depicts the sequence of actions needed to cluster or separate tickets into groups by similarity.

Clustering tickets in the repository

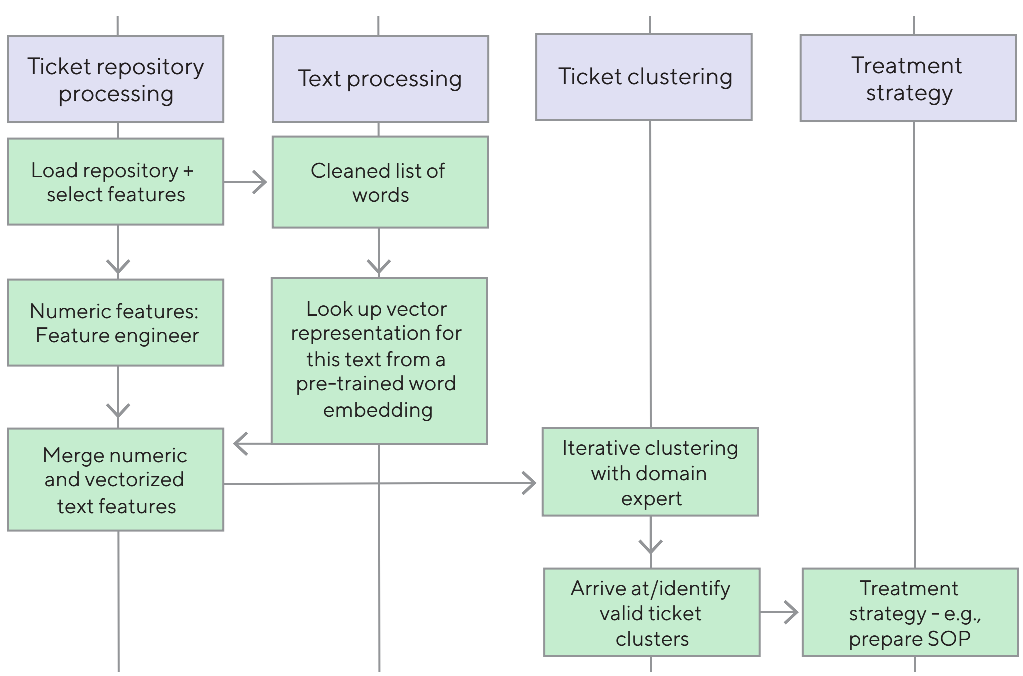


Figure 3

Assigning new tickets to a cluster

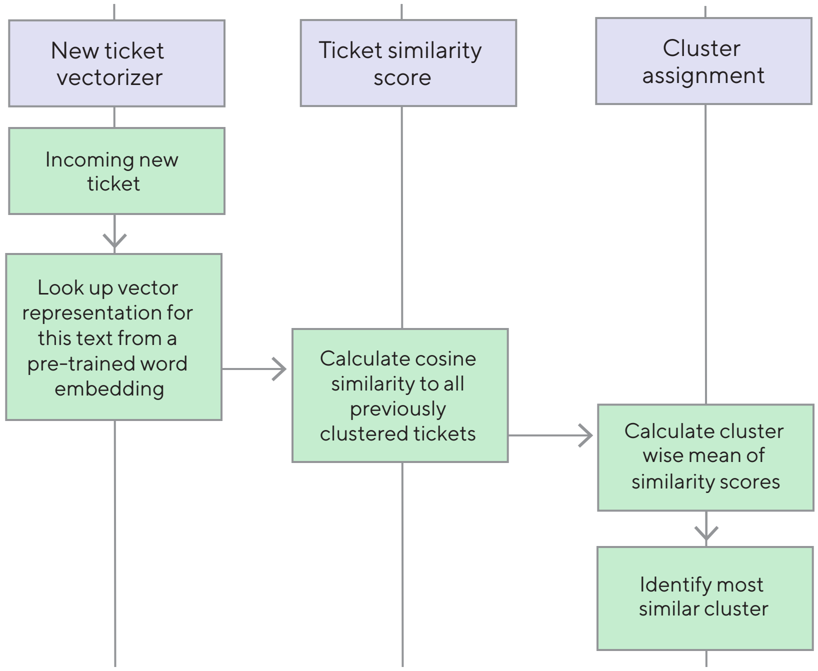
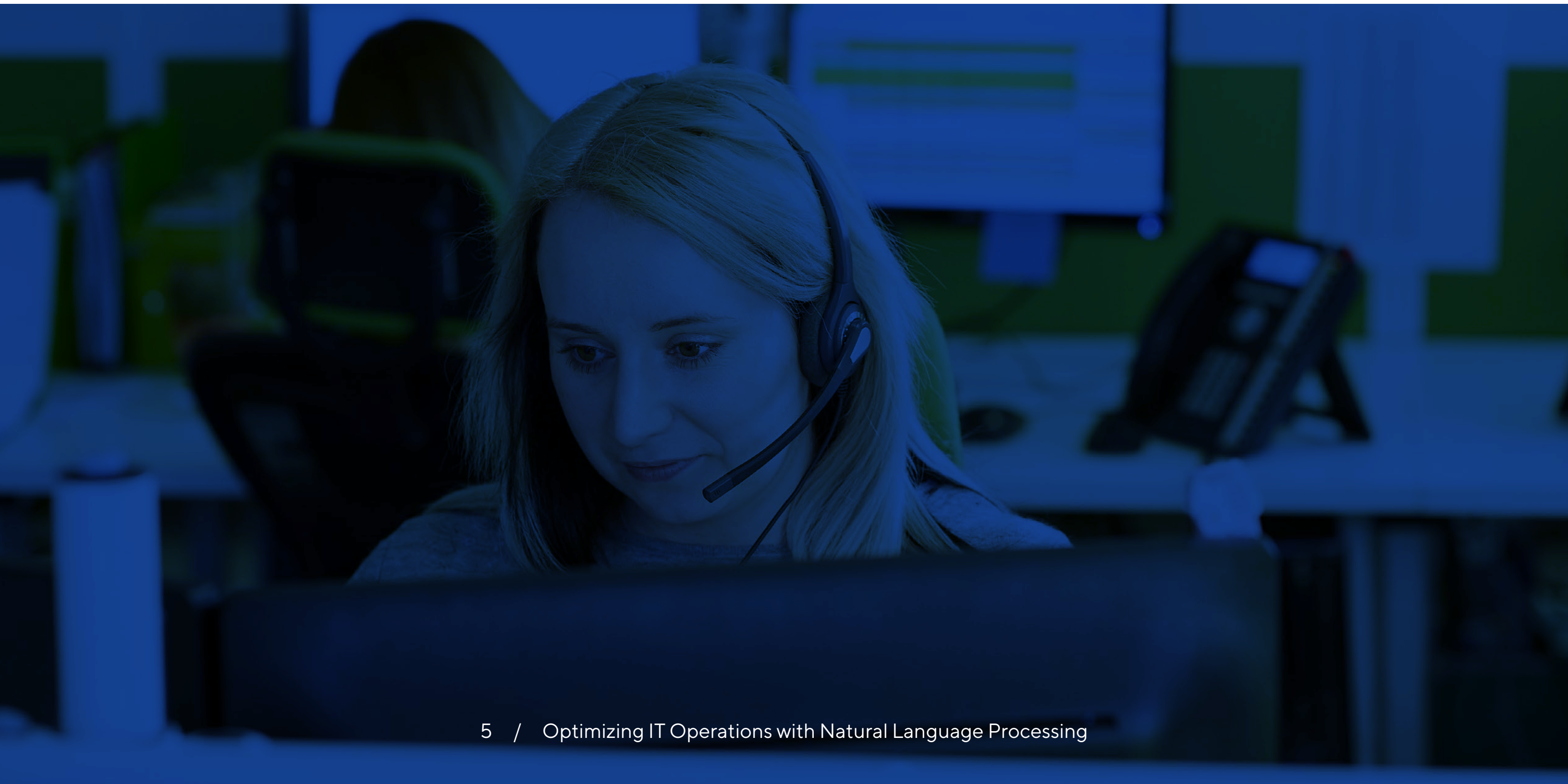


Figure 4

Figure 4 illustrates the process of assigning a new ticket to one of the previously calculated clusters.

The output includes a similarity score calculated using cosine similarity.[6]



Our solution is built in Python, using open-source libraries such as NLTK, BeautifulSoup, scikit-learn and Gensim (word2vec implementation), as depicted in Figure 5.

Our core solution applies across engagements. However, it typically requires iterations in which the SMEs fine-tune the hyper-parameters for the clustering algorithm. This solution is also more scalable and easier to maintain than the keyword-based alternate approach that many IT departments use. This automated approach is particularly useful for engagements with such large ticket volumes that it makes manual analysis unfeasible.

Our solution architecture

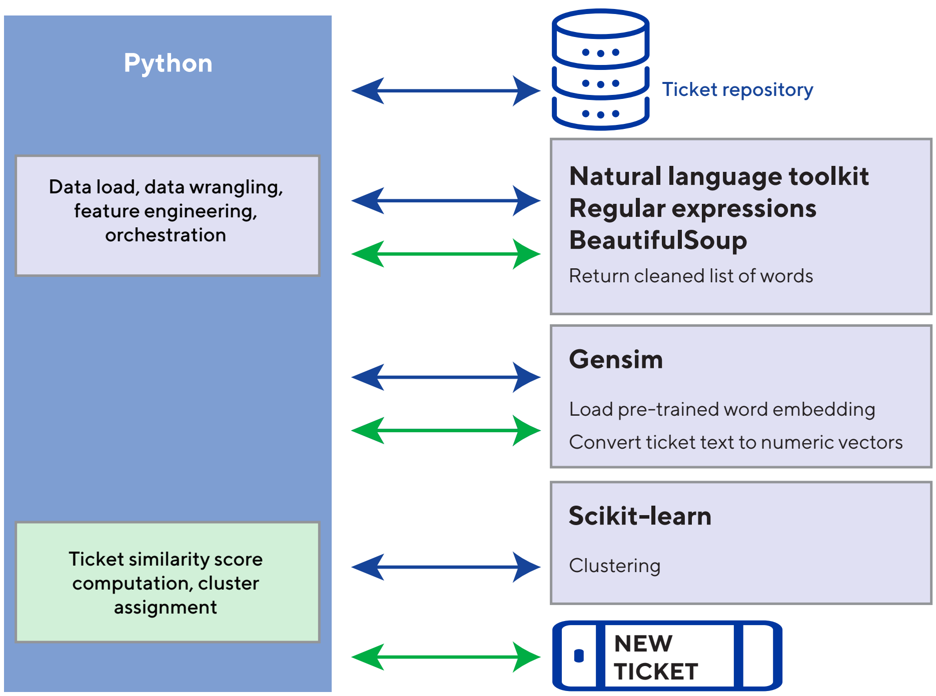


Figure 5

The business benefits are multifold. Our approach frees up IT operations team members who could be deployed on other tasks. It allows the deployment of less skilled and lower-cost resources to ticket handling tasks. It also enables quicker, more consistent and higher-quality ticket resolutions, which result in improved service levels and client satisfaction. Efficiencies will vary based on the nature of the engagement, depending on its inherent ticket patterns.

The solution also enabled the client to identify automation and problem management opportunities across the organization, which will eventually lead to nonlinear cost savings.

The opportunities ahead

Many opportunities arise from the ability to process free text ticket descriptions automatically without SMEs going over each one laboriously or identifying sets of keywords to categorize them.

We have piloted this solution successfully in engagements across industry segments. In an engagement with a life sciences client, benefits realized included:

- A 50% reduction in SME efforts for manual classification and identification of potential left shift and automation opportunities.
- Accelerating the benefit realization cycle by three months due to faster identification of opportunities.
- A 5% reduction in ticket counts and 20 percentage points improvement in first-level resolutions (FLR) in three months.
- A 20% reduction in mean time to resolve (MTTR) rates.

The solution also enabled the client to identify automation and problem management opportunities across the organization, which will eventually lead to nonlinear cost savings.

Other potential opportunities include deriving insights that will enable optimal team staffing and prioritizing knowledge transfer from an incumbent vendor when transitioning to a new partner. Application of the solution can be extended to other scenarios that require processing of free-form text, such as identifying patterns in software defects, identifying duplicate incident alerts, etc.

Endnotes

[1] Word embedding: "Vector space models (VSMs) represent (embed) words in a continuous vector space where semantically similar words are mapped to nearby points ('are embedded nearby each other')," https://www.tensorflow.org/tutorials/representation/word2vec.

[2] Word2vec: "Word2vec is a particularly computationally-efficient predictive model for learning word embeddings from raw text," https://www.tensorflow.org/tutorials/representation/word2vec.

[3] Clustering: "Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters)," https://en.wikipedia.org/wiki/Cluster_analysis.

[4] DBSCAN: Density-Based Spatial Clustering of Applications with Noise. This is a reliable clustering algorithm, originally proposed in 1996 and still widely used. It finds core samples of high density and expands clusters from them https://scikit-learn.org/stable/modules/clustering.html#dbscan.

[5] Silhouette coefficient: The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation); https://en.wikipedia.org/wiki/Silhouette_(clustering).

[6] Cosine similarity: Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them; https://en.wikipedia.org/wiki/Cosine_similarity.

About the authors

Bala Kesavan

Data Scientist, Predictive Analytics, Delivery Excellence, Digital Systems & Technology

Bala Kesavan is a Data Scientist, Predictive Analytics, Delivery Excellence, within Cognizant's Digital Systems & Technology line of service where he specializes in natural language processing. He focuses on tracking this continuously evolving space and applying innovative approaches to solve new problems and to improve current solutions. Apart from IT projects, Bala has domain experience in manufacturing and banking in a career spanning 20-plus years. He has a PGDM from MDI, Gurgaon, and can be reached at BalakrishnaSaravanan.Kesavan@cognizant.com | LinkedIn: www.linkedin.com/in/scmguru/.

Akhil Goyal

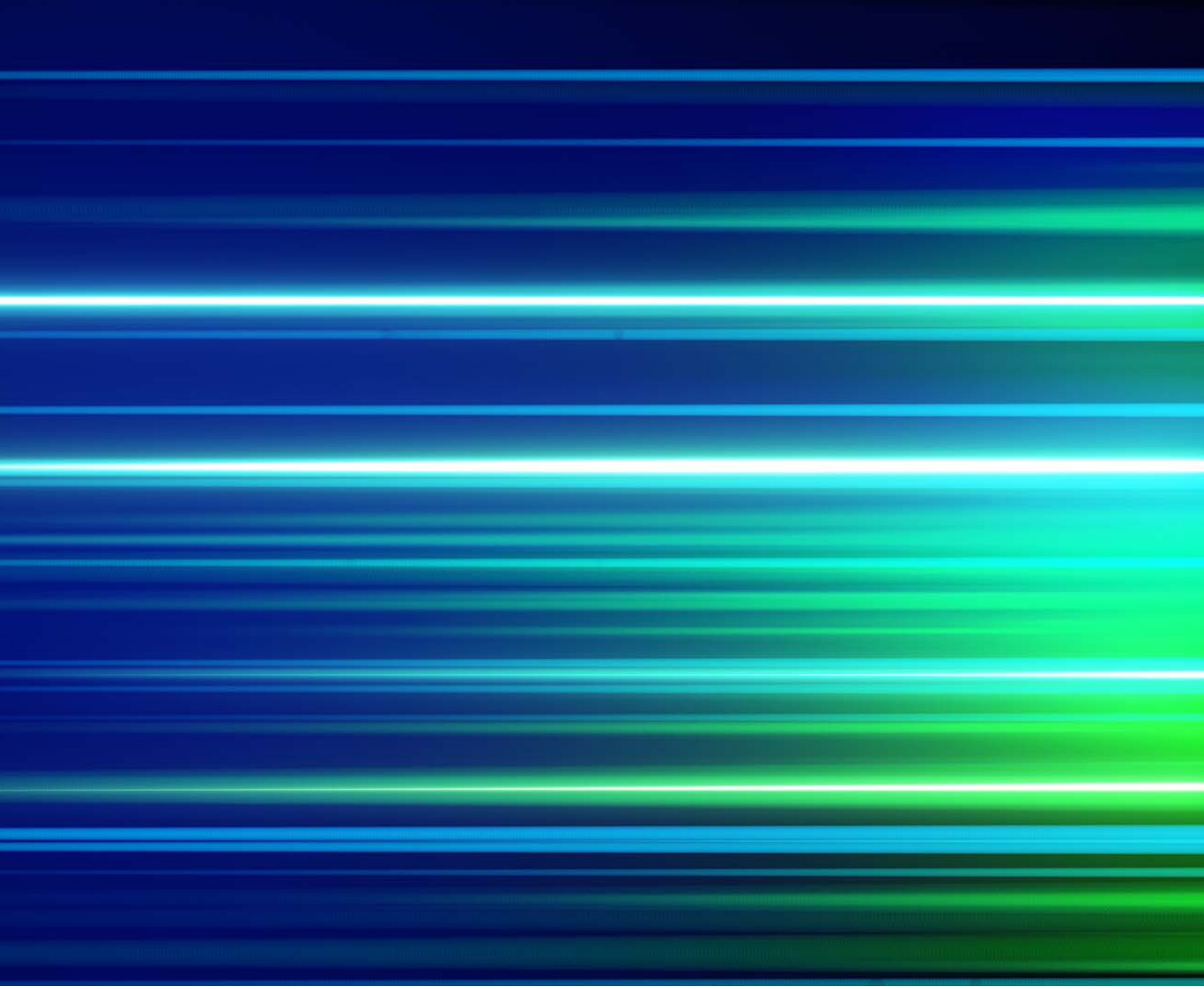
Director, Predictive Analytics, Delivery Excellence, Digital Systems & Technology

Akhil Goyal is a Director, Predictive Analytics, Delivery Excellence, within Cognizant's Digital Systems & Technology line of service where he leads the design and development of advanced analytics solutions across the company's delivery excellence initiatives such as knowledge management and reuse, risk management and contract management. He has over 20 years of experience in the IT industry across delivery management, client engagement, risk consulting and data science. Akhil has led multiple programs for global Fortune 500 clients and successfully deployed large organizational change initiatives. He received a B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi. Akhil can be reached at Akhil.Goyal@cognizant.com | LinkedIn: www.linkedin.com/in/akhilgoyal1.

Ravishankar Ganesan

Vice President, Delivery Excellence, Digital Systems & Technology

Ravishankar Ganesan is Vice President, Delivery Excellence, within Cognizant's Digital Systems & Technology line of service, where his responsibilities include strengthening pursuit robustness through market intelligence, industry benchmarks, risk assessments, solution and estimation review. He is also responsible for driving improvement in managed services revenue and deploying advanced analytics solutions to solve business problems. Ravi has over 30 years of diverse experience in quality assurance and delivery excellence and is a certified Black Belt and a certified assessor on the Malcolm Baldrige National Quality Award (MBNQA). He holds a bachelor's degree in mechanical engineering from Annamalai University, Tamil Nadu; a postgraduate diploma in statistical quality control and operations research from the Indian Statistical Institute; and an MBA from Indira Gandhi National Open University (IGNOU). He can be reached at Ravishankar.Ganesan@cognizant.com | LinkedIn: www.linkedin.com/in/ravishankar-ganesan-43309823/.



About Cognizant

Cognizant (Nasdaq-100: CTSH) is one of the world's leading professional services companies, transforming clients' business, operating and technology models for the digital era. Our unique industry-based, consultative approach helps clients envision, build and run more innovative and efficient businesses. Headquartered in the U.S., Cognizant is ranked 193 on the Fortune 500 and is consistently listed among the most admired companies in the world. Learn how Cognizant helps clients lead with digital at www.cognizant.com or follow us @Cognizant.

Cognizant

World Headquarters
500 Frank W. Burr Blvd.
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277

European Headquarters
1 Kingdom Street
Paddington Central
London W2 6BD England
Phone: +44 (0) 20 7297 7600
Fax: +44 (0) 20 7121 0102

India Operations Headquarters
#5/535 Old Mahabalipuram Road
Okkiyam Pettai, Thoraipakkam
Chennai, 600 096 India
Phone: +91 (0) 44 4209 6000
Fax: +91 (0) 44 4209 6060

APAC Headquarters
1 Changi Business Park Crescent,
Plaza 8@CBP # 07-04/05/06,
Tower A, Singapore 486025
Phone: + 65 6812 4051
Fax: + 65 6324 4051

Codex 4914 .02