



Unified Approach to Extracting, Normalizing and Analyzing Unstructured Data

Executive Summary

By most estimates, up to 90% of enterprise data is unstructured and never analyzed (Marr, 2019); 70% is in the form of text (e.g., emails, documents). By the end of 2021, more than 80% of organizations will fail to develop a consolidated data security policy across silos, leading to potential non-compliance, security breaches, and financial liabilities. (Davis, 2019) 71% of enterprises struggle with managing and protecting unstructured data (Rizkallah, 2017), and leading Robotic Process Automation vendors estimate that ~50% of their workflows still begin with a document. In addition, many organizations fail to leverage their data to produce insights into their processes, leading to potential loss of revenue or efficiencies.

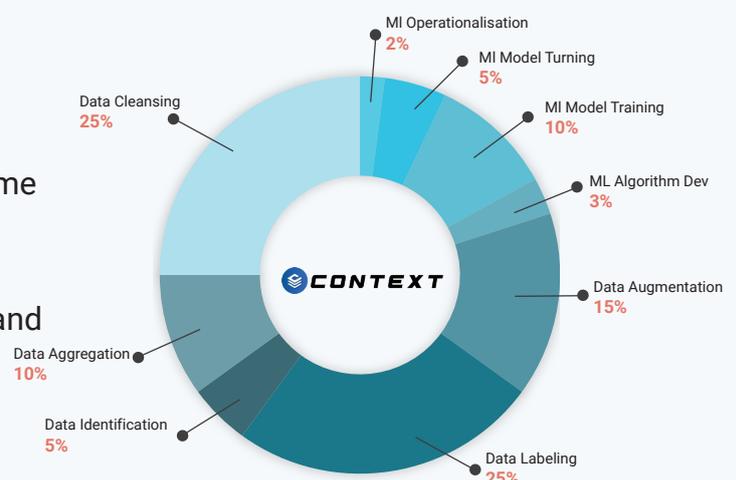
Business and Technical Challenges

There are few turn-key, out-of-the-box solutions that organizations can use to solve their problems. Most tools require custom development and API integrations that take months to implement and may not lead to a well-integrated solution. Once the data is extracted, organizing and normalizing different data structures can be daunting for most organizations and lead to cost overruns and delays. Data analysis can be challenging because the data is not organized in a single place, leading to siloed Analytics.

As of 2020, 80% of AI/ML projects have failed for reasons that can be summarized into 4 points (MV, 2020):

- 1) Challenging to get AI Projects off the ground
- 2) Non-AI data tasks take the longest
- 3) Integration between technologies
- 4) Custom Development is time-consuming
- 5) Adoption of data science is slow and cumbersome
 - a) Lack of experts
 - b) Slow Development time
- 6) Cost constraints due to long development time and complex pricing
- 7) Cloud Constraints – Different cloud providers provide different capabilities

Percentage of Time Allocation Machine Learning Project Tasks



Solution: Context - Take Complexity out of AI!

Our team at Areté has developed a fully integrated AI platform called Context that takes the complexity out of AI. Context gives organizations control over their data and empowers their workforces so that anyone can work with AI without understanding code. Our user-centered design, along with industry-leading AI models, allows organizations to upload, organize and analyze their unstructured datasets without any need to develop code or AI/ML models.



Vision AI

- Image Classification
- Face Detection
- Landmark Recognition



Video AI

- Object Detection
- Text Detection
- Explicit Content Detection



Intelligent Document Processing (IDP)

- Optical Character Recognition
- Document Classification
- Text Extraction



Web Archiving

- Archive any website available on the web
- Explore content using our intuitive UI
- Elastic search archived content



Speech-to-Text AI

- Speech-to-Text accurately punctuates transcriptions
- Upload your own voice data and have it transcribed
- Receive real-time speech recognition results



Tabular Data

- Data Comparison Utility
- Delimited, Excel and PDF file ingestion
- Automated PDF to table extraction

ContextAI Foundational Supporting Capabilities

Context's AI capabilities are supported by underlying features which allow organizations to gain insights into data, automate workflows, Perform advanced search, and improve user experience.



Flow Builder

Configure custom business rules and build advanced workflows using the Context Flow Builder



Enterprise Knowledge Graph

This exciting feature helps in entity enrichment, entity linkages and normalization of the extracted data from the documents.



Data Analytics

Make Data Driven decisions by using the advanced Analytics capabilities built into Context. Use our pre-configured analytics or build your own Analytics



Search

Our advanced search feature allows users to search for keywords and values throughout their datasets, whether structured or unstructured.



Human in the Loop (HITL)

HITL is an assistive tool that allows customers to introduce human reviewers in the digitization process where necessary, verifying and validating document metadata extracted by Document AI.



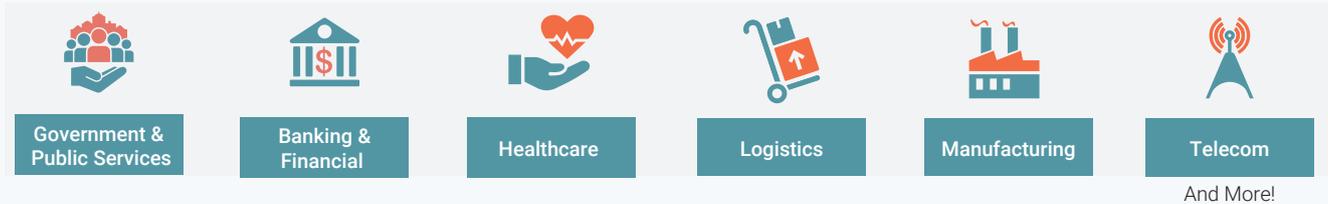
Data Warehouse

As part of the extraction, Context inserts the document metadata inserted into a Data Warehouse where it can be enriched with other sources and aligned with data governance initiatives.

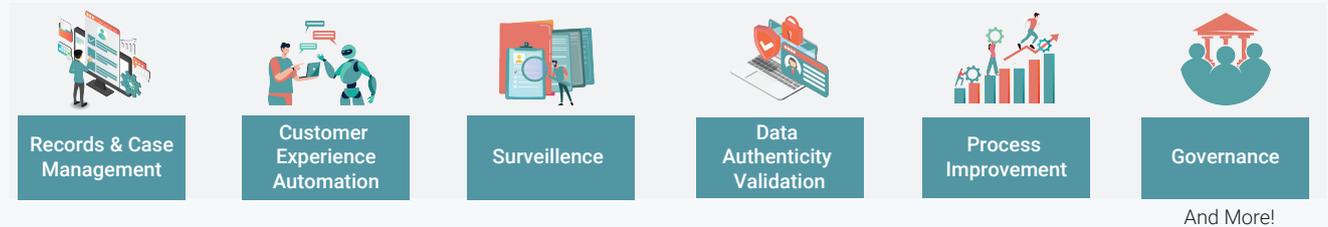
Industries & Use Cases

Start your Context journey to accelerate automation for distinct industries and use cases.

Industries



Use Cases



Aretec's Context Accelerators

- 1) Heuristic / Fuzzy Matching – Match key/value pairs using custom models
- 2) Custom Models – Our data scientists can build custom models for requirements not met by out-of-the-box models
- 3) Labeling-as-a-Service – Data Labeling and Annotation services for images, videos, documents, forms, and other niche use cases.
- 4) Model Uprating – Uprating of existing models for more accuracy
- 5) Workflow Automation – Custom automation using RPA + Context
- 6) Custom Integrations – Integrate with leading providers such as ServiceNow, Salesforce, AWS, Azure, Dropbox, Postgres, Oracle, and many others

About Aretec

Aretec, Inc. (Aretec) is a data science firm specializing in application development, quantitative research, and tailored advanced analytics platforms that transform data into actionable insights to aid federal government decision-makers. Aretec specializes in developing creative solutions to technical problems using appraised Capability Maturity Model Integration for Development Level 3 (CMMI-DEV/3) and CMMI for Services Level 3 (CMMI-SVC/3) engineering practices focused on customer requirements. Aretec is also appraised for ISO 9001:2015, ISO/IEC 20000-1:2011, and ISO/IEC 27001:2013.

Bibliography:

Davis, D. (2019). AI Unleashes the Power of Unstructured Data. Retrieved 28 July 2021, from <https://www.cio.com/article/3406806/ai-unleashes-the-power-of-unstructured-data.html>

Marr, B. (2019). What Is Unstructured Data And Why Is It So Important To Businesses? An Easy Explanation For Anyone. Retrieved 28 July 2021, from <https://www.forbes.com/sites/bernard-marr/2019/10/16/what-is-unstructured-data-and-why-is-it-so-important-to-businesses-an-easy-explanation-for-anyone/?sh=1e91500015f6>

MV, P. (2020). Top 10 Reasons Why 87% of Machine Learning Projects Fail - DZone AI. Retrieved 28 July 2021, from <https://dzone.com/articles/top-10-reasons-why-87-of-the-machine-learning-proj>

Rizkallah, J. (2017). Council Post: The Big (Unstructured) Data Problem. Retrieved 28 July 2021, from <https://www.forbes.com/sites/forbestechcouncil/2017/06/05/the-big-unstructured-data-problem/?sh=72e0ad51493a>

 **Headquarters:**
10201 Fairfax Boulevard, Suite 223 Fairfax, VA 22030

 **Orlando Office:**
300 S Orange Ave, Suite 1000, Orlando, FL 32801

 www.aretecinc.com

 info@aretecinc.com