Harnessing insights in unstructured data through NLP and ML

Beyond financial and transaction data

Financial and transaction data, also known as structured data, is commonly used for insights, analytics and reporting in financial institutions. On the other hand, unstructured data, such as call transcripts and emails, is largely an untapped resource due to accessibility and processing challenges.

The combination of natural language processing (NLP) and machine learning (ML) can offer powerful results. NLP is a machine's ability to interpret and analyze human language, both spoken and written, allowing us to extract context, meaning and intent from unstructured data. This capability, in combination with ML, provides an avenue to derive signals and generate actionable insights with a level of precision and scale that is significantly higher than ever before.

As the financial services industry continues to expand its analytics capabilities to leverage this data, new insights and signals from harnessing unstructured data can position financial institutions to better mitigate regulatory risk, drive growth via better customer experiences, and improve and design operations.

The evolving tech and analytics landscape

NLP and/or ML approaches are now more technologically and financially viable thanks to general advancements in computing power and processing capabilities, the availability of open-source software within analytics environments and the advent of new techniques. Additionally, technological developments have made turning speech into text easier and more accurate, allowing servicing data, such as customer call recordings and other audio-based data, to be candidates for NLP and/or ML solutions. Integration of optical character recognition (OCR) – which has seen significant improvements over the past decade – is also beneficial in unlocking previously inaccessible written data.

Applying these improved techniques to a variety of unstructured text data sources – such as emails, outbound marketing material, internal memos, chat transcripts, complaint logs and legal documents – is an effective way to enhance and automate regulatory review. It also provides an opportunity to mine insights to drive business improvement in other areas, such as service agent performance, customer sentiment and feedback, issue root causes, etc.

These approaches go far beyond rudimentary keyword searches or manually created lexicons that are commonly used in industry today. Open-source NLP repositories are much more sophisticated and act as accelerators in the development process. NLP techniques capture the ambiguities, variations and nuances of natural language, helping us identify relevant signals more comprehensively while reducing errors that make key-word-based techniques unwieldy and error prone.

Putting into practice

Consider how this can work on customer service calls: signals derived through NLP and ML can reveal trends and produce insights that can be used to identify customer disengagement or delinquency risk, call-handling issues and skills gaps among agents, and regulatory exposure with high levels of precision. These insights can be coupled with other data sources – such as customer transaction history, website and mobile app data, and other digital touch points – to provide a holistic view of the customer journey, identify cross-sell and up-sell opportunities and beyond.

Unlike human efforts related to call listening or manual review, NLP modules developed for pre-processing of unstructured data are extremely scalable, and financial services companies can rapidly expand these solutions to a broader set of custom use cases.

Proper deployment of NLP/ML solutions to create new, impactful capabilities requires cross-disciplinary teams encompassing advanced data science experience and business domain expertise. Developing an approach that mimics human efforts is a guided process – it's not simply developing algorithms, it requires learning and incorporating the human decision-making process.

Domain experts and data scientists should work together to adequately define the use-case requirements and the problem scope – two areas that are often overlooked when adopting new technology. An effective solution is a balance between allowing machines to detect patterns and guiding the development work with domain expertise.

Data-driven solutions should be embedded within existing workflows so that they are easily adopted as incremental enhancements, rather than disruptions. This can take form in custom dashboards and user interfaces, to visualize trends, facilitate and expedite manual review processes, and provide ML solutions with feedback loops for improved precision and coverage in the future.
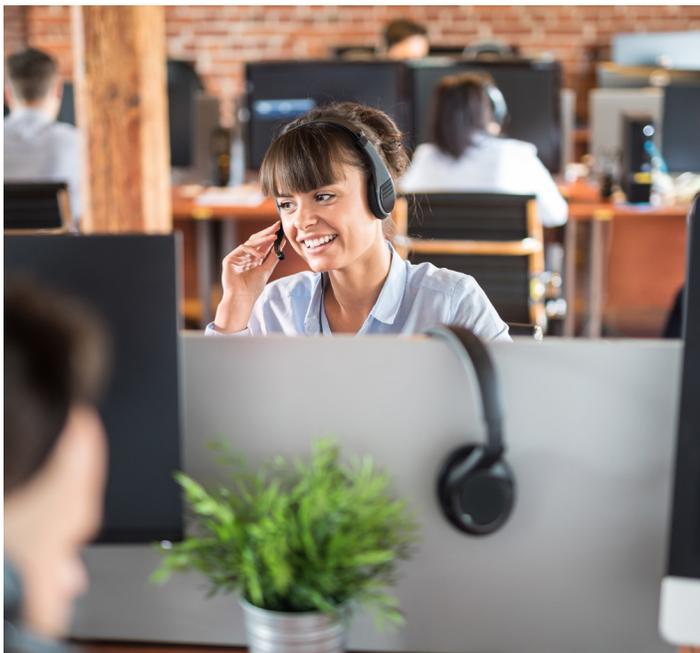
Considerations for successful implementation

To successfully implement these advanced capabilities and capitalize on the business impact, organizations must consider these areas:

1 Your current processes

To find potential areas of improvement, pinpoint which processes rely heavily on manual efforts and where the appropriate technology and unstructured data are not being leveraged. Which areas have the highest scope for automating processes and reducing cost?

2 Your current state

Where do you stand in terms of adoption of cutting-edge technology/ML approaches? Financial services companies must examine if they are maximizing the capabilities available and determine where they stand against other peers in their industry.

3 Your starting point

What are some areas where you have accessible data and can begin to establish advanced solutions? Organizations should consider which use cases are opportunities to make rapid gains and will ultimately expedite their adoption process.

4 Your infrastructure

Developing NLP and ML models requires a computing environment able to process throughput and access to a broad suite of NLP libraries. This is crucial for scaling your approach and leveraging it for a broader range of use cases.

5 Your talent

Do you have access to the right talent and resources? Developing cross-disciplinary teams with the right mix of data science and industry expertise is a necessary step to succeed. External resources can also help gain perspective and stay well-informed on emerging trends and industry leading practices.

Looking ahead: Unstructured data no longer has to be a blind spot, left for cost-intensive and inefficient manual reviews. Now, NLP and ML open the door to a host of insights that protect the business, reduce costs and support growth.

1

Automated sales practice identification for increased compliance efficiency and regulatory coverage

A bank received more than 500,000 customer complaints annually. Its existing process for identifying sales practice issues relied heavily on key-word filtering in customer call transcripts with a manually crafted lexicon – and then a team of hundreds of employees performed manual monitoring, review and validation. How could the bank apply NLP and ML for more effectiveness and efficiency?

The EY approach

Recorded customer calls were converted into transcripts. Extensive pre-processing of the transcribed calls decoded misspellings, acronyms and jargon in the text. An Ernst & Young LLP (EY) team layered multiple NLP techniques – including semantic pattern matching and sentiment analysis – to identify and extract customer signals. Those signals informed the development of linear and nonlinear ML models to predict the risk of sales practice issues.

The outcome

By using the EY approach, the bank achieved a more than 90% capture rate of sales practice issues, reducing the risk of regulatory matters requiring (immediate) attention. False positives declined by more than 30%, and manual reviews dropped 80%, so this application of NLP also drove increased operational efficiency.

2

NLP/ML assisted workflow for marketing material review

A financial institution needed to review its marketing materials to determine whether they were in breach of fiduciary limitations set by a recent Department of Labor (DoL) ruling. The client's existing setup relied completely on manual review by the compliance team. How could the process be automated, considering that the marketing materials were in multiple formats and document types and included a broad range of terms?

The EY approach

The EY team developed a proprietary tool to parse and review the materials by deploying ML and NLP algorithms, which were packaged into a virtual assistant to aid in the overall review. EY also developed a module that extracted information based on the document layouts. The tool was designed to fit naturally into the institution's existing workflow. The approach identified the presence of disclosures and glossary terms, flagged promissory and misleading language, and calculated the degree to which documents adhered to DoL requirements laid out in templates.

The outcome

The EY approach reduced the review time per document by 95%, and it did more than just minimize the resource requirements. It also accelerated the process of reviewing outliers and consistently captured standard issues with improved consistency across document types, which relieved regulatory pressure and reduced time to market.

Contacts

Sameer Gupta

AI and Advanced Analytics Leader

Ernst & Young LLP
sameer.gupta@ey.com

Darrin Williams

AI and Advanced Analytics Leader

Ernst & Young LLP
darrin.williams@ey.com

Contributors

Sukumar Pant

Ernst & Young LLP
sukumar.pant@ey.com

Talia Visaggi

Ernst & Young LLP
talia.s.visaggi@ey.com

Visit us at:

ey.com/fsadvancedanalytics

EY | Assurance | Tax | Transactions | Advisory

ey.com