ADVANCED ANALYTICS AND DATA MINING

Identifying Rare Geopolitical Events Early With News and Derived Data

Geopolitical Risks Becoming More Important

Geopolitical events with a large impact on companies and institutions are rare, tail risk events by nature. However, since they might impact considerably, for instance, international trade, it is important for companies and investors to spot such rare events early on and to be able to react accordingly. Examples include regulatory, national security or economic policy changes that impact the terms and balance of trade across major geographical regions.

Company executives and investors need to have reliable means and processes to identify and assess such risks as early as possible. This requires an automated, machine-based approach that relies on high quality data sources such as Dow Jones DNA. With more than one billion articles from 8,600 sources, as well as DNA Streams providing financial and business news in near real-time, DNA proves to be the right foundation for advanced geopolitical data mining and analytics efforts.

Topic Modeling and Rare Event Identification Automated

Modern data science, natural language processing (NLP) and machine learning (ML) approaches allow for the automated and efficient screening of large amounts of data in real-time. Via automatic topic modeling approaches—for instance, combining NLP techniques for word embeddings with unsupervised ML techniques for clustering—rare geopolitical events can be spotted almost as soon as they happen. They can also be followed along to monitor their evolution and impact on businesses, regions and risk profiles.

DNA provides standard APIs to query historical data, as well as to retrieve data in near real-time. News articles, for example, are delivered not only in machine-readable formats, but also come with comprehensive meta data—event-type classifications and mappings to relevant political bodies. This allows anomaly models to flag rare, tail risk events automatically. This makes DNA the ideal basis for automated topic modeling and the identification of rare events.

How to Use Dow Jones DNA for Automatic Rare Event Detection

The head of a strategic risk analysis group, which provides risk management and consulting services to an internal client group, wants to improve their geopolitical risk identification process. So far, it is mostly based on manual work of political analysts and media monitors responsible for different global regions. The business unit head plans to use DNA to automate the risk identification process as far as possible. To this end, they hire an external data scientist who has worked with DNA before. Together they fix a set of key building requirements for the application.

Guided by the requirements, the data scientist writes code that queries DNA's API at regular intervals to collect the most recent news articles related to the predefined set of topics. Using DNA, they also create a taxonomy of major stakeholders for every key topic. Based on this corpus of topics, data and stakeholders, the data scientist implements algorithms that categorize the articles retrieved, summarizes the single articles, models topics automatically and identifies events of interest. In addition, articles are tagged which present comments of important stakeholders on a topic.

After the successful implementation of a broader test program based on historical data, the application is deployed for near real-time monitoring. During the first week of deployment, overnight, a rare event is detected—a severe data breach by a large European retail financial institution. The analyst responsible for the region is notified automatically, first thing in the morning about the event. They screen the related articles—retrieved from DNA and summarized by the application—and assesses the severity of the event. Later in the morning, the automated screening process identifies and publishes first comments by government bodies.

Based on the provided information and the different reactions by stakeholders, the analyst assesses the impact and consequences the event might have. They discuss these with the responsible political analyst. The analyst compiles a brief note summarizing the background of the event and the reaction of the governmental stakeholders. The note discusses the impact the event might have on the regulation of retail financial institutions in Europe and also, on related data protection laws. At 12:15 pm, the note is sent out to executive stakeholders that are active in this industry or that might be affected by the consequences of the event—for example, institutional investors holding shares in retail banks.

News data for advanced and augmented analytics

Technology and advanced BI tools have created new opportunities for organizations to rapidly gain intelligence and create opportunities. News as an alternative data source can support advanced analytics initiatives such as:

- Data/text mining
- Machine learning and NLP
- Forecasting and predictive analytics
- Pattern matching and trend spotting
- Visualizations
- Semantic and sentiment analysis
- Network and cluster analysis including graph analysis
- Complex event processing

Dow Jones DNA - Data, News & Analytics

At a time when data fuels the professional world, Dow Jones DNA gives you data for AI and allows you to seamlessly connect datasets. One of the world's most comprehensive licensed news datasets, DNA is designed to readily integrate with your organizations' advanced analytics in order to provide deep insights and automate business decisions.

DNA is a cloud based Data-as-a-Service platform to help you leverage outside insights and increase the accuracy of your data outputs. You can:

- Have confidence and reduce risk with news and data you can trust, from our 31+ year archive of proprietary and licensed news data with storage rights through contract life
- Rely on highly veracious data with 8,600+ sources in 28 languages from extensive regions, industries and topics
- Save time and increase productivity with well-structured metadata from our cleaned and labelled datasets. Features include tagged company codes on 20m+ companies and standardized formatting of timestamps across 1.3bn articles
- Scale and tailor the specifications and delivery method to best suit your data science teams. Our DNA Solutions Engineers are here to assist you integrate DNA

What will you build with DNA?

To learn more, visit **www.dowjones.com/dna** or contact your sales representative.

If you are located in the US, you may also call **800-369-0166**.