Monetizing Information Consumers on a Financial Research Platform

MSBA Practicum Proposal submitted by BlueMatrix

Created July 29, 2020

About BlueMatrix

Founded in 1999, BlueMatrix is a SaaS company which provides software for the authoring, distribution, and analysis of the consumption of financial research. These research reports are a quantified analysis of the state of a company, sector, or region and an analyst's opinion on that topic.

BlueMatrix's 120 clients, mostly small and mid-size broker/dealers and independent research firms, author 500,000 research reports each year. BlueMatrix sends the reports to major aggregators in bulk and to individuals in over 500 million emails annually. Additionally, over 1,000 independent research providers who author content on their own platforms (typically on Microsoft Word) use BlueMatrix to distribute them.

In total, over 60% of all research written in the United States is either created on or is distributed by BlueMatrix. BlueMatrix is also making inroads in Europe; This week, Extel Surveys[1] identified the top four research providers in Germany, each of whom use BlueMatrix to author and distribute content.

Financial Research Industry Background

In the late 1990s, the mechanics of authoring and distributing research involved manually entering content into aggregator platforms (e.g. Bloomberg.) It was an exceedingly tedious, time-consuming and error-prone process destined only to grow in complexity as regulators required more legal and compliance-related disclosure and disclaimer information.

It was around the turn of the millennium when, capitalizing on the rapid rise of the web, technologists began to develop tools to more efficiently and accurately publish research. Until then, technology had advanced at a slow pace, limiting research analysts' ability to gather and distribute quality work and effectively putting those firms that could not pay large numbers of staff at a disadvantage.

[1] https://www.extelsurveys.com/

The largest banks responded by investing large sums into developing internal solutions intended to amplify their advantage in the market. Lacking the same resources, smaller banks and boutique operations turned to third-party vendors that had begun to launch affordable tools designed to streamline the authoring and distribution process.

It was these smaller operations and otherwise unheralded analysts who benefited from the new technologies the most. Armed with the ability to showcase their expertise, broaden their readership and expand stock coverage, these providers could, for the first time, compete with the largest Wall Street players on output, despite having fewer resources.

Very quickly, a new marketplace unfolded, one where analysts could produce more research and do it more accurately, therefore freeing up time to focus on higher-value activities, like content.

Fast forward to the situation following the financial crisis of 2008-9. To help meet the renewed demand for stock research, vendors focused on breathing new life into research with more engaging and powerful delivery methods. Platforms put readership into investors' hands with pull versus push distribution platforms, and analysts enticed their audience with dynamic, responsive formats, like HTML5 appropriate for mobile devices. These delivery methods not only came with sophisticated modelling capabilities, but also let analysts track which – and how much of – reports their customers consume, allowing research departments to more efficiently allocate time and resources.

Another round of technical innovation came after the January 2018 introduction of MiFID II. MiFID II is a legislative framework instituted by the European Union (EU) to regulate financial markets in the bloc and improve protections for investors. Its aim is to standardize practices across the EU and restore confidence in the industry, especially after the 2008 financial crisis.

Innovation has focused on features that track interactions or protect intellectual property by ensuring only entitled customers are accessing their research. These tools help providers concentrate and direct their research efforts and dollars by identifying less actionable or poorly reviewed work, bringing new levels of efficiency to the research creation process.

Today, new technology trends stand to further drive innovation. Using artificial intelligence and machine-learning tools, analysts can read enormous amounts of presorted and pre-processed data quickly to transform them into reports. Other sophisticated investors are using natural-language-processing tools to mine market sentiment from multiple analysts' reports on the same stock and plug it into trading engines that determine where, when, and how to place their orders.

A new application sure to make its mark is the distributed ledger framework, often referred to as blockchain. By its nature, blockchain enhances transparency and mitigates risk, making it a powerful tool for highly regulated industries like financial services.

All in all, technology has helped streamline how equity analysts have produced and distributed research over the past 20 years, much as regulations have reshaped the industry's structure and reined in its excesses and abuses.

The investment research industry is more important than ever due to the tremendous and unprecedented uncertainty in the world today due to the global pandemic. Investors, decision makers and the general public are craving information on a scale not seen before. The internet and maturity of the web makes access to content easier than ever. However, since the content is produced from highly regulated entities, mature well thought out platforms are essential to guarantee the authenticity of the content and the fair and appropriate dissemination of the content.

BlueMatrix is well positioned to meet the needs of the investment research industry with over 20 years' experience, a large share of the worlds content, and a staff of highly trained technologists. More information on BlueMatrix and the research industry is available on its web page[2].

Strategic Opportunity

As research is consumed, BlueMatrix captures thousands of data points on the consumption of this content: emails received, times spent on pages, reads on aggregators (BBG, FactSet,) time spent in certain components (e.g. mouse hover of a chart, or table) etc.

Today, BlueMatrix provides uncorrelated basic statistics to its customers regarding the consumption of their research. For instance, they can learn that seven people read this document at 9 AM, one hundred at 10 AM, one thousand at 11 AM, etc. The topic of the documents read was on Electric Cars, PPE, etc.

Modern methods for interpreting and associating data may indicate where and how our customers compare to their peers, possibly correlated to market factors. To gauge the effectiveness of an analyst, more advanced analysis could compare the number of reads relative to peers or relative to subsequent market changes. BlueMatrix has done this in a rudimentary fashion in the past, building one off spreadsheets and pivot tables based on static relational queries. There is an extraordinary opportunity to leverage modern data analysis techniques for more precise, timely, meaningful, and monetizable solutions.

Project Overview

BlueMatrix services two main customer types, the "Sellside" producers of research and the "Buyside" consumers of research. Historically, Sellside analysts emailed research reports to clients, and to aggregators such as Bloomberg or FactSet. Today, the trend is to send emails

[2] https://www.bluematrix.com/

with links to PDFs or HTML5 based content on their own portals. Another emerging trend is to provide additional audio and video explaining and summarizing the research.

Historically a “read”, considered an indication of interest, was captured by a primitive pixel embedded in each email. Today, this is becoming a less and less valuable indication of interest due to the forwarding of emails, email (virus) scanners, and automated consumption scripts. Using HTML5 or PDF documents rendered online using more sophisticated viewers, BlueMatrix can capture time spent on a page, mouse locations on a page, the read of a document from disparate IP addresses. These provide a more accurate indication of interest.

Goals

This collaboration with UC Davis has significant potential to benefit BlueMatrix *and* its Sellside and Buyside customers in numerous ways. Early on in the project, BlueMatrix will work with the student team to create a formal Statement of Work containing a sequence of phased sub-projects. As the engagement progresses, the SOW may be amended with the guidance of project staff and concurrence of the student team.

The rest of this section contains an ordered list of scoped opportunities to create business value through analytics. In each case, the minimum goal is a thorough analysis and description of the data required to solve the problem. Ideally, the participants will also produce a prototype, at least in Excel, to demonstrate the data processing, analysis and modeling required to produce a solution usable by the company.

Semantic/NLP Document Search/RIXML Analysis

BlueMatrix provides portals to its Sellside customers in which they showcase research to their clients. BlueMatrix also provides an aggregated platform to its Buyside customers in which they can search for and view the research from all firms to which they are entitled.

Today, BlueMatrix uses a very basic Lucine[3] solution to execute simple queries. A more modern system or method to provide search capabilities is desired, e.g., using Scala[4]. The results of the search would certainly go beyond basic string matches to associated words, or possibly even sentiment-based associations.

Readership Benchmarking

Aid the BlueMatrix sales and marketing team by normalizing client activities across the 1100 different research producers. Though there is a normalized schema for research, RIXML, the vast variety of classifications in use make it difficult to accurately aggregate information.

[3] https://lucene.apache.org/core/
[4] https://towardsdatascience.com/simple-nlp-search-in-your-application-step-by-step-guide-in-scala-22ca1ce3e475

For instance, one research firm may refer to the ticker APPL, another the company name “Apple Inc,” a third by the CUISP 037833100, and yet another by the CIK 320193. Modern data science techniques could help BlueMatrix determine, for instance, which analysts writes the most popular research on Apple, finally making a true Apple to Apple comparison**!**

BlueMatrix could then rank our clients on readership by sector, company, region, country. The purpose here is to sell this as a service to our clients (a bit like the McLagan[5] rankings) and even entice the Bulge Bracket to join.

Predictive behavior based on history

In the three to four weeks following the last month of each calendar quarter (December, March, June, and September) most public companies release both revenue and profit results from the quarter just ended. During these four “earning seasons,” BlueMatrix observes a significant increase in the volume research produced and consumed.

BlueMatrix has had a hard time comparing previous earnings seasons due to an inconsistency in the number of clients serviced, the number of analysts at each firm, and changes to the sectors, industries, or companies those analysts cover. With better predictions on what may happen, BlueMatrix could better prepare for higher volumes prior to certain companies reporting or notes of certain nature being released.

Modern Tool Advice

BlueMatrix plans to formalize its Machine Learning capabilities through the selection of modern tools ranging from TensorFlow to Watson. The design and development of *a prototype* on these or similar tools would help advance that effort.

Requirements

The practicum team will be expected to:

1. Utilize common tools such as MySQL and (S)FTP to view and access BlueMatrix data.
2. Provide a detailed ontological analysis of BlueMatrix data and, working with the BlueMatrix team, identify meaningful associations which may not seem obvious.
3. Produce prototype systems (e.g., in Excel) that reliably provide results with a measurable business impact.
4. In conjunction with the BlueMatrix team, develop specifications for (possibly enterprise-grade) systems to be developed.

Items 1-2 are expected to be delivered by mid-January, 2021. The application specification in Item 3 is expected to be finalized by mid-April. The contours and timetable for item 4 will largely depend on the disposition of preceding sub-projects, but at any rate the associated product specifications have been delivered and achieved company buy-in by the end of May, 2021.

[5] https://mclagan.aon.com/business-performance/banking-capital-markets-wallet-share

Data

BlueMatrix will provide access to multiple terabytes of data including

- RIXML data
- Research report textual content
- Obfuscated readership data (relational) primitively mapped to attributes
- Access to Sellside and buyside customers where appropriate

Skills and Technologies

Previous familiarity with, or the ability to quickly come up to speed on, the financial sector, and especially financial research, will be helpful. Excel skills are key to interacting with Sales and Marketing staff.

Published document data are manifested as RIXML (metadata) and in PDF or HTML5 format. The data for readership is stored in common relational databases. Direct access could be given to Linux (CentOS) environments to run native SQL queries. (Administrator can build out machines with arbitrary resources and tool kits, e.g. python libs, R, etc.)

SQL, Python, JSON and XML formats, and ability to call web server APIs are sufficient for most of the work. Core applications are built in Java/JavaScript. A working knowledge and ability to code in these languages would be a huge plus for more advanced scenarios.

Resources

BlueMatrix will provide access to appropriate (remote) environments for the student team to access databases and any other IT infrastructure. BlueMatrix will furnish any required devices (virtual machines,) tools, reasonable licenses and access to third party systems (e.g. AWS) needed to perform the appropriate analysis.

BlueMatrix primarily uses Office 365 for email and business productivity tools, and Slack and Zoom for communication. BlueMatrix will make available its library of in-house training videos and documentation through its Atlassian Confluence instance. The student team may be directed to utilize our Atlassian Jira instance for project management.

The student team will primarily interact with our top two engineers working in this space, Shaun Codner, and Casey Mullins.

Shaun is a Senior Developer at BlueMatrix with over 20 years of experience in the financial and technology industries. Prior to joining BlueMatrix in 2016, Shaun worked at Goldman Sachs, Bell Laboratories, and IBM. In his current role, Shaun handles the collection and analysis of the large volumes of data generated by BlueMatrix business clients. Shaun has an MS in Computer Science from Cornell University and a BS in Computer Science from New York University.

Casey Mullins is a Software Engineer at BlueMatrix with close to 10 years in the financial technology industry. Prior to joining BlueMatrix in 2015, Casey performed R&D at BEST Technologies (a global macro hedge fund). In her current role, Casey works closely with Shaun collecting, transforming, and analyzing large amounts of data generated by our BlueMatrix clients. Casey has an MS in Chemical Physics from the University of California San Diego and a BS in Physics from Davidson College.

Efrain Rivas, one of BlueMatrix's North America System Administrators, will provide access to machines, tools, licenses, and be available during regular business hours to help facilitate any system-based troubleshooting, such as, trouble logging in, exporting data, etc.

CIO Daniel Faltyn will interact with the team on a regular basis, coordinating the overall project direction and facilitating meetings with the sales and marketing team and select clients. He is very interested in not only working with but also learning from today's student thought leaders.

Daniel joined BlueMatrix in 2014 after 5 years at JP Morgan and 10 years at Goldman Sachs where we was awarded multiple patents for his work in mobile device security. Daniel holds bachelor's degrees in Mathematics and Computer Science from the University of Scranton, and a master's in computer science from Columbia University. Daniel started off his career at DuPont on a technology AI team writing expert systems.