



Predictive Analytics for Institutional Investors – Success Story

Disruptive technology solution for deeper insights into performance of prospective companies

Application: Financial Performance Database and Dashboard

Services Offered: Financial Data Harvesting from public sources of information. Scraping automation. Database Design and data storage. Building a dashboard product to convert the information to insights.

Tools: Python, AWS RDS, Tableau

Client

The client is a Financial Services & Product startup that provides financial consulting services to customers such as institutional investors, private equity investors, mutual fund managers. They are building a financial product that provides aid on industries and stocks with respect to their performance, comparison with the peer stocks, industry revenue trends etc. The product is designed to be customized for different kind of audience like retail investors, institutional investors to give analytical insights for decision making on stock purchases.

Key Highlights

Key Success:

Product to scrape multiple formats of data

Automaton to scrap data and handle exceptions

Domain:

Financial Services

Duration:

4 months and ongoing

QA Team:

2 developers, 1 architect, 1 manager

The Requirement

A team of three, realizing the key role data analytics can play in helping institutional investors assess potential companies for investment, started a fintech venture to develop a technology solution for corporate data analytics. The solution was expected to analyse core company data to project potential growth in the coming few quarters, considering various internal and external factors influencing this, and thereby giving investors the insights needed to assess the suitability of the venture for investment. This is a disruptive process as earlier auditors would personally examine the profit and loss statements and use their intuition to arrive at the decision, which was prone to human error.

The fintech company needed an analytics development partner for creating a solution that would take publicly available information about companies, analyze the past performance and project the future growth potential. Indium with more than two decades of experience in working on cutting edge technologies and the fintech space was found to be ideally suited and commissioned with the task.

The Challenge

While it is relatively easy to cull this information for listed companies, smaller ventures and unlisted companies do not have enough data in the public domain, especially revenue details, making this culling difficult.

What data is there is available in multiple formats in the public domain, including tables, pdfs, plain text, videos and images, which increase the burden of getting easy to archive and analyze financial information.

Progress

Indium is in the process of creating a product that would help retail investors, institutional investors and equity managers unearth insights about the companies' health and revenue. These insights are more than simple P&L info and other stock charts indicators. These insights compare actual vs projected revenue of companies aiding in informed decisions to investors.

The Indium Approach

Indium developed a three-pronged approach to meet the requirement as well as overcome the challenges of insufficient and unstructured data. Its solution includes Reusable Data Extraction, Flexible Data Architecture and Data Analytics capabilities.

Through daily scrapping, its tool extracts structured data easily accessible in the public domain. Where this is not possible, it used Power Law modeling technique – a statistical model where the functional relationship between two quantities is established such that a relative change in one quantity results in a proportional relative change in the other quantity: in other words, one quantity varies as a power of another.

While this data is stored in MySQL, the unstructured data is cleaned before storing. Then this data is analyzed and presented in charts created using Tableau and Qlikview for easily understanding the performance trends.

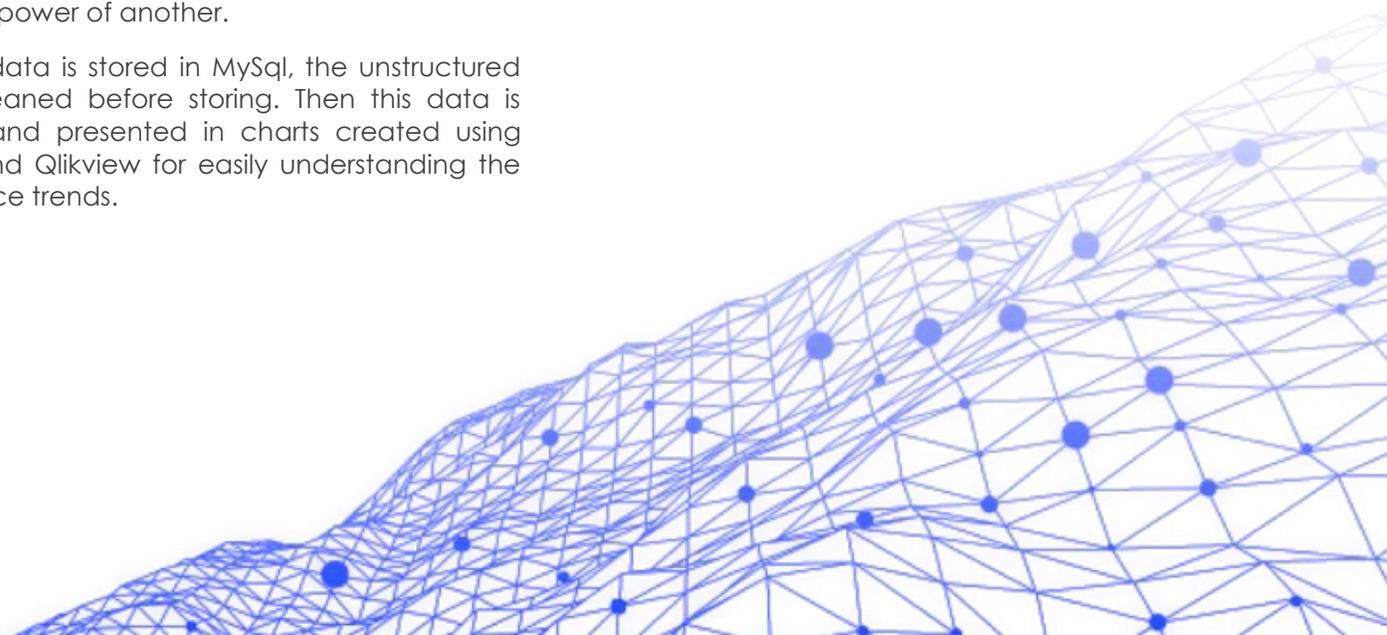
The Scraping Process

While data scraping is a simple process of extracting the data from the web and other publicly available resources, it becomes complex when one wants to get value from it. It requires

- Careful selection of data sources
- Writing customized code as per the data formats
- Meticulous cleaning of the data
- Transformation of unstructured to structured data
- And finally, storing in a database in an optimally queryable format.

This whole sequence becomes a multi-step project, informally known as data harvesting.

Indium processed data in some complex formats such as PDFs, unstructured plain text and table within tables. It also scheduled the routines so that the automaton wakes up every day and looks for updated data from multiple sources. There was logic embedded to trigger any errors and exceptions that was forwarded to the managers when data couldn't be extracted because of reasons such as irregular update at the source, change in format or system downtime issues. Now the system runs like clockwork extracting data in less than 10 minutes for 100+ sources without any manual intervention. The right ingredients are ready for the Sous Chef to cook.



Transformation and Engineering

Process

Simple scraping of data is not enough for full-fledged analysis as additional information needs to be trimmed and appended wherever there is a lack of information. Indium transformed parts of data and created new fields as part of analytics and stored it in a secured cloud infrastructure. This provided a solid data foundation for visualizing the trends of the past revenue and predicting the future revenue potential. With the visualization product as the final goal, Indium

- Optimized the fields in the databases to only the mandatory ones for speed and efficiency
- Optimized the derived field calculations
- Coded to perform the calculations during the non-peak hours.

Outcome

As a proof of concept, Indium's solution has demonstrated successfully the revenue and growth potential of four companies for the next quarter. This POC will be used by the fintech venture to showcase its product to external investors and fintech experts. They would assist the client in forging a bigger vision for the product and expanding the product to multiple stock groups. The client will go full steam with the development of the solution targeting different audience of retailers, Mutual fund house, Equity advisors, Institutional investors etc.

Impact

The techniques implemented by Indium helped the client in building an automated robust data collection routine.

Despite scraping from multiple data sources, the data was standardized. Unstructured data was cleaned for removing inconsistencies, unwanted noise resulting in organized SQL tables storage.

Using statistical techniques such as Power Law and Levenshtein Distance, data was crunched in such a way that more than 98% accuracy on prediction was achieved during the analysis.

The product is built in a scalable mode on Big Data technology as the calculations on time, storage were planned beforehand considering the vision of the client.